

农业病虫害远程诊断与预警技术

The Remote Diagnosis and Early Warning Technologies for Agriculture Disease and Insect

李道亮　傅泽田　温继文　段青玲　著

清华大学出版社

霍英东教育基金会高等院校青年教师优选资助项目

农业病虫害远程诊断与预警技术

Remote Diagnosis and Early Warning Technologies for Agricultural Disease and Insect

李道亮　傅泽田　温继文　段青玲　著

清华大学出版社
北　京

内 容 简 介

本书系统地总结了农业病虫害远程诊断和预警技术的相关概念、理论、方法、技术和应用，全书23章组织成相应的四篇：①理论篇。提出一套基于诊断理论、预警理论和会诊理论的能够指导农业病虫害远程诊断和预警的理论体系。②方法篇。系统地研究了农业病虫害诊断和预警知识的表示、推理和获取，为农业病虫害诊断问题以及处理疑难杂症问题提供有效的求解模型。③应用篇。结合网络技术、计算机技术、短信技术、移动PDA技术、视频会议技术和呼叫中心技术，总结了农业病虫害诊断领域的应用经验，特别是本书作者所在科研团队的长期探索，为相关人员提供技术支持。④示范篇。在上述研究成果的基础上，给出了农业病虫害远程诊断和预警专家系统的开发和应用示范。

本书适合农业病虫害诊断与预警相关领域的科研人员与工程技术人员阅读，也适合各级农业病虫害防治机构、农技站以及农业病虫害防治技术人员阅读。

图书在版编目（CIP）数据

农业病虫害远程诊断与预警技术 / 李道亮等著 .--北京：清华大学出版社，2010.12
ISBN 978-7-302-22928-5

Ⅰ. ①农…　Ⅱ. ①李…　Ⅲ. ①作物－病虫害－诊断　Ⅳ. ①S431

中国版本图书馆 CIP 数据核字(2010)第 101192 号

责任编辑：赵彤伟
责任校对：刘玉霞
责任印制：

出版发行：清华大学出版社　　**地　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　编**：100084
社 总 机：010-62770175　　**邮　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 刷 者：
装 订 者：
经　销：全国新华书店
开　本：185×260　**印　张**：18.25　**字　数**：459 千字
版　次：2010 年 12 月第 1 版　**印　次**：2010 年 12 月第 1 次印刷
印　数：1～0000
定　价：0.00 元

产品编号：　-01

序言

PREFACE

我国是一个农业病虫害多发的国家，主要农作物病虫害达1400余种，每年因病虫害造成粮食损失约5000万吨，棉花100多万吨，蔬菜、果树的病虫害更是严重。2007年全国农作物病虫害发生面积达到53亿亩次，严重发生面积3.5亿亩次，防治面积累计达到50多亿亩次。全国每年水产养殖病害发病率达50%以上，损失率20%左右，每年因水产养殖病害问题而造成的直接经济损失就达百亿元之巨，并且还有上升的趋势。病虫害在导致巨大经济损失的同时，也常造成环境污染与食物污染，对人类食品安全构成威胁。

我国农业病虫害专家比较缺乏，平均每万名农业人口中科技人员仅有6.6人，加上农民的科技素质还比较低，绝大多数农户还处在半文盲或文盲状态，对病虫害发生的规律认识不够，因此不能科学地防治病虫害，很难适应病虫害监测、管理及预测防治的要求。为此，如何准确、及时、快捷地监测与诊断病虫害，并提供病虫害的诊治方法，是有效控制病虫危害、保障农业安全的基础性工作。

为解决病虫害发生频繁而防治专家缺乏的矛盾，国内外许多专家学者对病虫害诊断与防治进行了研究，主要解决途径是研制开发小麦、玉米、蔬菜、棉花、果树病虫害专家系统以及畜、禽、鱼虾疾病诊断专家系统，利用这些专家系统模仿人类专家对病虫害进行诊断和防治。但是，由于计算机和网络在农村的普及程度还很低，加上从实际应用中获取知识的限制，以及农民文化素质低等原因，目前已经开发的病虫害诊断专家系统几乎为零，大大降低了农业专家系统应用效果。同时，病虫害一旦发生就必然造成不同程度的危害，因此预警技术显得尤其重要，在病虫害没有发生之前，就应对其进行合理预警，采取必要的预防措施，使病虫害发生控制在萌芽阶段。

随着现代通信技术和网络技术的发展，特别是呼叫中心和移动通信技术的发展，加上我国农村电话装机量(包括手机)的迅速普及，将短信技术、移动PDA技术、视频会议技术、呼叫中心技术、预警技术与Web专家系统技术集成，实现病虫害的远程诊断与预警已经成为可能，能从根本上解决专家缺乏、专家系统应用效果差等问题。

病虫害远程诊断与预警技术研究旨在根据症状、疾病、病因、治疗内在逻辑关系，提出基于短信技术、移动PDA技术、视频会议技术、呼叫中心的病虫害诊断数据的采集、挖掘和智能推理方法，探讨基于Web的专家系统与短信技术、移动PDA技术、视频会议技术、呼叫中心的合理集成方案，研制开发病虫害远程诊断与预警系统，实现普通坐席人员与病虫害领域专家对远距离之外的用户多种模式的诊断、咨询和预警，最大限度地解决我国病虫害严重、病虫害专家缺乏、农民素

质低等一系列现实问题。

病虫害远程诊断与预警技术研究先后得到国家高技术研究发展计划(“863”计划)“农业病虫害网络化远程诊断技术研究与示范”(合同号:2002AA243031),霍英东教育基金会高等院校青年教师优选资助项目“病虫害远程诊断技术”(合同号:94032),国家科技支撑计划“淡水健康养殖网络管理与疾病辅助诊治系统”(合同号:2006BAD10A02-05),国家“863”计划“集约化水产养殖数字化集成系统研究与应用”(合同号:2007AA10Z238)4 个课题的支持。

撰写本书的目的在于系统总结农业病虫害远程诊断和预警技术相关概念、理论、方法、技术和应用,并突出以下四方面的工作。①理论体系:提出一套基于诊断理论、预警理论和会诊理论的能够指导农业病虫害远程诊断和预警的理论体系,为病虫害诊断与预警研究提供理论方法;②方法模型:系统研究农业病虫害诊断和预警的知识表示、知识推理、知识获取以及群决策模型,探索获取农业病虫害诊断知识,准确、真实表达农业病虫害知识,有效求解农业病虫害诊断问题以及处理疑难杂症问题的解决模型,为相关人员提供方法支持;③技术支持:结合网络技术、计算机技术和数据分析技术、短信技术、移动 PDA 技术、视频会议技术、呼叫中心技术、预警技术与 Web 专家系统技术,系统总结国内外,特别是著者所在研究团队在农业病虫害诊断领域的应用探索经验,为相关人员提供技术支持;④示范和应用:在上述研究成果的基础上,结合软件工程实现农业病虫害诊断和预警专家系统的开发和示范应用。从理论、方法、技术、系统、应用五个层次与读者分享我们的研究工作。

本书共分 4 篇。第一篇为理论篇,论述用于指导农业病虫害远程诊断和预警实践的理论,主要包括农业病虫害诊断理论、预警理论、群决策的会诊理论、专家系统理论。第二篇为方法篇,论述近几年本研究组成员在农业病虫害远程诊断和预警领域构建的方法模型,主要包括农业病虫害的知识表示模型、基于节约覆盖集的鱼病诊断模型、基于禁忌搜索算法的鱼病诊断模型、基于遗传算法的鱼病案例推理模型、基于案例检索策略的鱼病诊断模型、基于 CBR 和 RBR 协同推理鱼病诊断、基于相对熵的集结偏好模型、基于证据理论的 FDT 群决策方法、基于遗传算法的知识获取模型、基于粗糙神经的知识获取模型。第三篇为应用篇,重点讲述了我们科研团队先后开发的系统:基于 J2EE 的鱼病诊断系统、基于案例推理的鱼病诊断系统、基于呼叫中心鱼病诊断专家系统、移动式鱼病诊断专家系统、基于 Web 的奶牛病害诊断系统、基于 Web 的果树病害诊断系统、集约化牙鲆疾病预警系统以及重大动物疫情疫病咨询系统等。第四篇为示范篇,重点讲述了病虫害远程诊断技术的示范和应用。

山东农业大学王衍安博士撰写了“基于 Web 的果树病害诊断系统”一章,本科研团队中的研究生周云、郭永洪、杨萍、戎立斌、张杰、周小燕、祝伟、胡利平、李丽、邢斌做了大量富有成效的研究工作,这里一并表示感谢!

由于作者水平有限,书中的错误或不妥之处在所难免,诚恳希望同行和读者批评指正,以便今后改正和完善。

作　者

2010 年 6 月 1 日于中国农业大学

目　录

CONTENTS

理　论　篇



方 法 篇



应　用　篇





示 范 篇

理 论 篇

第1章

绪　论

1.1　概述

我国是一个地域广阔、地理条件复杂、人口众多的农业大国,经济相对落后,由于承灾能力较差、生态基础脆弱等因素的综合影响,农业生产经常遭受灾害的侵扰。据统计,我国常见的农业有害生物有1648种,而且分布广、突发性强、成灾频率高,每年都有几种重大病、虫、草、鼠害暴发和流行,造成每年损失粮食数十亿公斤、棉花300万～400万担、木材近千万立方米,再加上水果、蔬菜、油料、水产及其他损失,每年的总损失近百亿元。农业生物灾害在导致农业减产和歉收的同时,由于防治不当等原因,也造成了农产品质量的严重下降。随着生态环境的进一步恶化,农业生物灾害发生的频率和造成的损失也逐渐加大。因此,农业生物灾害的可持续控制是关系到农业可持续发展以及我国食品安全的重大问题,重大生物性灾害可持续控制的基础理论研究是农业生产的急需,且具有重大的科学意义。

近十年来,随着国家的支持和投入的增加,我国的农业生物性灾害研究工作逐步加强,取得了许多重要成果。从全国范围来看,已初步建立了县、乡(镇)植保公司、植物(庄稼)医院、防治专业队等生物防治体系,专业化防治农作物病、虫、草、鼠害面积占防治面积的10%以上;并且已经初步建立起各类灾害监测预警系统,提高了预防灾害的能力。科研院所也在国家支持下,对农业生物灾害的控制进行了一系列研究,取得了一系列成果。中国农业大学中欧农业信息技术研究中心也是其中的一支队伍。

但与发达国家相比,我国目前农业科研投入低(仅为国家年度财政预算总金额的0.17%～0.27%,而世界平均值为1%,发达国家超过5%),并且由于缺乏一套完善的农业科技推广与合作运行机制,不能使推广的职能得以有效发挥,在某种程度上存在着产学研脱节现象,进而导致农业生物灾害监测、预报和防治工作中存在着很多问题,总体来讲科技含量偏低,而部分较先进的监测、预报系统适用范围小且功能单一,监测预报准确率不高,甚至还有许多植保科研成果尚未在农业生物灾害的防灾减灾中发挥作用。同时,由于缺乏常规性和规范化的技术标准,无法形成技术整合平台,许多成功的经验和信息不能共享,造成重复建设以及技术之间的矛盾冲突。

因此,为了及时准确地掌握农业灾害发生和发展状况,预防和降低各种变化的不利影响和灾害的破坏程度,并促进农业生物灾害可持续控制技术的共享,利用数字化、网络化、智能化等先进的现代信息技术,对农业生物灾害可持续控制技术进行集成、战略重组和系统优化,形成能够对重大农业生物灾害监测和防治技术进行研发、共享与推广平台是非常必要的,并且具有长远的实质性意义。

本书的研究成果具有可视化、智能化、构件化的特点,所包含的农业病虫害知识领域广泛、

信息量大、操作简单,可延伸性、可维护性和安全性强、在线直观等特点。该系统的推广与示范为实现农业生产发展由数量型向质量型转变,提升农业病虫害防治的技术与管理水平,降低农民生产成本,提高农民收入,促进我国农业信息化发展打下了良好的基础。

本书是对近十年内研究组成员在农业病虫害诊断和预警领域取得的成果和实际应用的深入汇总和全面完善。

1.2 诊断理论及其应用

1.2.1 诊断推理概念的提出

"诊断"意指鉴别、确定。推理就是根据一个或几个已知事实,找出事实之间蕴涵的关系,从而推导得出新的事实的过程。[1]在医学领域,诊断推理就是对已知症状或发现的结果利用有关的知识推导出为什么会出现这些症状或结果的解释。其过程实质上是一种"假设—测试"的循环推理过程:首先根据诊断问题的初始信息,产生一组似然假设,这些假设可以解释初始信息,然后对这些假设进行测试,并指导获取更多的信息,直到产生最优解释作为诊断问题的解。[2]

诊断推理来源于反绎推理,是反绎推理在诊断领域的应用。反绎推理与演绎推理、归纳推理是人类逻辑推理的三种基本方式。有关研究可以追溯到古希腊时代,当时亚里士多德就认识到存在一类不同于演绎的推理过程,如"IF P THEN Q "当 Q 为真时,结论可能是 P 或其他,显然,结论不是 Q 的逻辑推演,而是一种猜测或假设。这类推理一直被作为归纳推理的一部分而加以研究,直到美国哲学家 C. S. Peirce 首次将这类推理命名为"abduction",即反绎推理。[3]

20 世纪 70 年代反绎推理的研究开始面向诊断问题,形成了"诊断推理"这一概念。同时,Pople 开发了第一个运用诊断推理的系统 INTERNIST-1,此后又出现了 Neurologist 系统、PIP 系统、IDT 系统,但缺乏一般化和形式化的研究。20 世纪 80 年代中后期,出现了一般性的诊断推理模型和诊断推理的形式化研究工作,如 Reggia 提出了集合覆盖的思想并建立了节约覆盖集理论。20 世纪 80 年代末 90 年代初,Yun Peng 等人在进一步发展节约覆盖集理论的基础上,提出概率因果模型,实现了符号推理与概率推理的结合,但不能解决现实中模糊性诊断问题,随后 Yager 提出模糊解释模型,但没有给出有效算法,Josephone 基于诊断推理机制提出了最优解释假设集成器。20 世纪 90 年代初,杨叔子等人提出模糊覆盖集诊断模型,实现了符号推理与模糊推理的结合,并应用于故障诊断领域。[4]

1.2.2 诊断推理的方法

人工智能技术如专家系统、知识工程、数据挖掘等的快速发展,为诊断技术提供了新的更加有效的方法及问题求解思路,并促使诊断技术向着"智能化"、"知识化"阶段迈进。按照不同的建模方法和处理手段的性质及特点,诊断方法可分为三大类:基于解析模型的方法、基于信号处理的方法以及基于知识的方法。[4]

基于解析模型的诊断方法通过利用一些数学、物理学上的准确模型对检测目标的状态进行检测和评估来实现诊断。该方法的特点就是诊断的准确率很高,算法在计算机上实现比较简单,但要求检测目标必须具有比较精确的数理模型的特点,因而限制了该方法的应用



范围。

基于信号处理的诊断方法直接利用相关函数、频谱分析、小波分析等信号模型来进行诊断识别,它回避了建立或抽取诊断对象的数学模型的技术难点,使该方法有较好的适应性,尤其是小波分析方法,对噪声的抑制能力强,对故障检测有较高的灵敏度。但该类方法的诊断准确率要比基于解析模型的方法低,主要适应于线性系统。

基于知识的诊断方法是以知识处理技术为基础,以推理为主要手段来实现推理过程与算法的统一、辩证逻辑与数理逻辑的集成、符号处理与数值处理的统一、知识库与数据库的交互与协调调度。该方法研究的是基于知识的诊断问题求解过程中所涉及的主要理论、技术、方法以及实现等问题,包括概念体系、诊断知识处理与应用、诊断系统的结构与实现等方面的内容。该类方法不需要定量的数学模型,是一种很有前途的方法,是智能故障诊断技术的核心。基于知识的诊断方法包括规则推理方法、模糊推理方法、模式识别方法、神经网络方法和定性模型方法等。[5,6]

1.2.3　诊断型专家系统

诊断型专家系统的任务是根据观察到的数据(事实)推出相应诊断对象存在的故障(或机理失常),找出相应的原因并给出解决方案的一类专家系统。诊断型专家系统是开发、应用最早,也是目前应用最多的一类专家系统,主要包括医疗诊断和机器故障诊断专家系统。其中医疗诊断专家系统又广泛应用于人类医疗、动物医疗、植物医疗等领域。[7]

20 世纪 70 年代末期,美国伊利诺伊大学 R. S. Michalski 等人推出 PLANT/ds(大豆病害诊断专家系统);A. G. Boulanger 等人开发了 PLANT/cd(预测玉米黑地老虎危害的专家系统);美国弗吉尼亚工学院 J. W. Roach 等人开发了 POMME(农业害虫和果树管理专家系统)。20 世纪 80 年代中期,日本千叶大学园艺部开发了番茄病害诊断专家系统 MICCS。

国内在农作物病害诊断方面,中国农业大学以及中国农业科学研究院等农业研究机构分别研制开发了“农作物病害诊断系统”;南京农业大学与安徽农业科学院联合设计开发了“水稻主要病害诊断防治专家系统”;哈尔滨工业大学研制开发了“玉米病虫害诊断专家系统”和“大豆病虫害诊断专家系统”;重庆大学研制开发了“番茄栽培病虫害诊断专家系统”;河北农业大学先后研制开发了“面向果园的苹果、梨病虫害防治决策支持系统”和“基于 GIS 的农作物病虫害防治决策支持系统”。在动物疾病诊治方面,国内先后开发了鸡、猪、牛、羊等畜禽病诊断专家系统以及鱼病诊断专家系统等。[8~16]

1.3　会诊理论及其应用

1.3.1　群决策理论和方法

随着科学技术和生产的发展,社会、经济和生活中的许多决策问题变得越来越复杂,单个决策者很难考虑问题的所有方面。为减少决策的失误,很多企业和组织都要雇佣许多人来共同研究决定,这就是群决策。从单个决策者背景向多个决策者背景的转变使决策分析变得更加复杂。

对群决策问题的研究始于 200 多年前,法国数学家 Borda 于 1781 年提出了群体对方法排序 Borda 规则;1785 年,法国另一位数学家 Condorcet 提出了 Condorcet 规则,并发现了投票

悖论;1944 年 Von Neumann 和 Morgenstern 对多人对策问题中的效用函数进行了研究;1951 年,诺贝尔奖获得者、美国经济学家 Arrow 提出了著名的不可能定理,在一组看似非常合理的公理和条件下,并不存在集结社会中各成员偏爱的社会福利函数,这一结果为群决策奠定了重要的理论基础,并对社会的政治和经济都产生了深远的影响。[17~19]

由于群决策问题的内在复杂性,群决策理论既是决策理论的前沿,也是决策理论最为薄弱的部分。目前群决策理论和方法的研究还很零散,尚未形成一定的框架体系,所使用的理论仍是效用理论和对策论。但总体来讲,国内外学者对群决策的研究主要集中在两个方面:一个是行为集结,即过程导向型方法,侧重于对群决策行为、过程等的研究。该类方法着眼于决策的动态过程。运筹学和管理科学的相互作用过程及组织心理学方法,都可以看成是过程导向型方法。另一个是数量集结,即内容导向型方法,侧重于如何更好地让决策者给出偏好信息,以及对群体中决策者偏好信息如何进行集结并得到群的偏好等方面的研究,如群体效用理论、模糊决策理论等。第一种方法论或多或少地可以看成是基于效用理论的内容导向方法的动态过程,第二种方法着眼于决策者行为的演化和在群决策过程中适当的磋商技巧的使用。最典型的就是德尔菲法和 NGT 法,此外还有交互式群决策方法等。群决策理论的研究还存在着很多问题,群决策在实践中的应用也还需要进一步研究。此外,鉴于群决策理论研究的主要是静态的偏好集结模型,而实际上群决策是一个信息反复交流最终达成一致的动态过程,所以加强对群决策过程的研究是十分必要的。[20]

20 世纪 90 年代,随着计算机技术、网络通信技术的发展,为消除或减少决策个体之间信息交流的障碍提供了可能,群决策的绩效也得到了较大的改善。我国对群决策理论的研究是从 20 世纪 80 年代开始的,从国内对群决策进行的研究来看,主要是对群决策数学模型方法、群决策支持系统(GDSS)以及社会选择理论的研究等。[21]

1.3.2 群决策支持系统

20 世纪 80 年代末以来,DSS 研究与运用的热门话题之一就是群决策支持系统(GDSS)。所谓群体决策支持系统是相对个人决策支持系统而言的,将两人或多人召集在一起,讨论实质性问题,提出或设计解决某一问题的若干方案、评价这些方案各自的优劣,最后作出决策,这样的决策过程可称为群体决策。与 GDSS 相关的概念还有电子会议系统(EMS)、计算机支持协同工作(CSCM)、群件(groupware)等系统。

GDSS 研制应用的目的就是要提高决策会议的质量和效率。现实中一直存在着会议次数多、时间长,却难以解决参与者意见不易充分表达、群体智慧难以生成而影响决策质量与效率的问题。人们希望通过 GDSS 式的电子会议改善群体决策,使之不仅不受时间、地点的限制,且能通过信息技术和社会人文知识的应用,形成防止消极群体性格和行为发展的内部机制,如更好地将参与者的注意力集中于相应的问题;帮助理解复杂的问题与环境;阻止不当信息和群体思想扩展;更好地促进群体成员相互交流、了解和信任,避免有害的抑制和冲突等。[21]

群决策希望解决的问题是集结一群人中每个人的偏好以形成群的偏好,然后根据群的偏好对集结方案进行排队,从中选择群所最爱好的方案。因此,群体必须选择一种能寻求一致意见的方法,以便找到最优解或满意解。

GDSS 这一概念包括“决策”、“支持”和“系统”三部分,因此从较高的层次看,GDSS 主要的研究内容应是:

(1) 群体决策。包括对个人和群体决策心理或认知过程的研究以及对群体交互社会学的



研究。

(2) 决策支持。即研究开发各种技术支持手段和特征,以方便群决策过程中信息的收集、显示、交换、管理和协调等活动。

(3) 支持系统。即设计和开发各种应用系统,以满足不同环境和组织的决策支持的需要。

1.3.3 远程医疗会诊系统

远程医学(telemedicine)是集计算机管理系统、现代通信和现代医学为一体的一种新型医疗服务项目。一个远程医学系统作为一个开放的分布式应用系统,应包括远程诊断、专家会诊、信息检索和远程教学等几个主要部分。它的主要优点是:病人可跨越时空地域的限制,选择远方的著名专家看病,真正做到快捷、方便、经济、高效,这就为合理配置和充分利用医疗资源,开辟了新领域和新途径。

20 世纪 50 年代末,美国学者 Wittson 首先将双向电视系统用于医疗;同年,Jutra 等人创立了远程放射医学。此后,美国相继不断有人利用通信和电子技术进行医学活动,并出现了“telemedicine”这一词汇,现在国内专家统一将其译为“远程医学”。[22]

自 20 世纪 80 年代后期,随着现代通信技术水平的不断提高,一大批有价值的项目相继启动,这代表了第二代远程医学,其声势和影响远远超过了第一代技术。远程医疗在国外的发展已有四十多年历史,目前已经发展到了第三代,其特征有:通过公众通信网提供服务——入口变得容易;提供瘦客户端到桌面的服务——维护变得简单;成熟项目商业化——动机变得单纯。从 MEDLINE 中所收录的文献数量看,1988—1997 年的 10 年间,远程医学方面的文献数量呈几何级数增长。在远程咨询、远程会诊、医学图像的远距离传输、远程会议和军事医学方面取得了较大进展。纵观国外远程医疗的发展史,可以发现它与电子技术的发展息息相关,经历了“电话时代”和“电视时代”,直到如今这个通信与计算机网络飞速发展的“多媒体时代”。[23]

1.4 预警理论及其应用

预警原是一军事术语,指对可能出现的敌情预先发出警报,后来预警逐渐发展为对某一警素现状和未来进行测度,预报不正常状态的时空范围和危害程度以提出防范措施。近年来,预警已在洪水预报、农业经济、气象、饥荒、疾病、地质灾害、环境等方面得到广泛应用,同时也渗透到农业病虫害领域。

1.4.1 病虫害预测预报

预警和预测之间既有区别又有联系,预测只能对未来状态的某一指标给出量化值,并不能表征病变的变化程度,预警恰好能解决这一问题,对未来变化趋势及好转或恶化的程度给出一个明确的概念。预测预报作为预警的前期阶段,主要经历了经验预测法、实验预测法、统计预测法、信息预测法和组合预测法五个发展阶段。[24,25]

(1) 经验预测法

由于预报组织的不健全,病虫害发生资料积累较少,并受到科学发展的制约,人们对病虫害的预测预报一般建立在经验基础上。经验预测法在国内外病虫害预测中被广泛使用,对指

导农业生产起着重要作用。它是通过一定的经验积累,明确病虫害的发生种类、数量和危害损失程度等定性关系,应用这些现象对病虫的发生期、发生量和灾害程度等进行估计的方法。经验法中所用简单相关系数的相关程度不高,做出的预测很难达到较高的准确率,同时由于预测时限短,不能满足人们制定长远规划的目的,因此经验预测法存在一定的局限性。

(2) 实验预测法

实验预测法是应用实验生物学的方法,研究探讨植物营养状况、气候条件、自然天敌、土壤环境等因素对病虫害的生存、传播或迁移、生长繁殖的自然规律,并利用当地的气候、土壤等资料对病虫害的发生和危害情况进行估计。实验法是病虫害预测的基础方法,常用于中短期病虫发生情况的定量预测。由于实验法利用一些影响病虫害发生发展的关键因子和必然规律进行病虫害发生的预测,其预测准确性较高、指导病虫害近期防治的实践性强,具有较强的应用性。

(3) 统计预测法

统计预测法是利用病虫害的发生危害情况与病源、虫源、气候、环境等因素存在的必然联系,根据多年或多点的病虫害调查和观测等积累的数据资料,采用相关分析、回归分析等数理统计方法进行适当的统计分析和计算,组建出各种预测模型或数学方程,通过预测模型估计病虫害的发生和危害情况。

统计预测法原理和方法比较简单,容易掌握。但统计预测法仍存在着两个不可忽视的问题,一个是预测效果不稳定,有时预测方程拟合率高,但实际应用时随条件变化而使预测准确率降低,即泛化能力不强;另一个是缺乏深厚的生物学和物理学基础,有些统计预测模型纯粹建立在相关普查基础上,没有或很少考虑病虫害的生物学和物理学规律,缺乏统计预测模型的生物学理论基础,并不能完全把病虫害的发生规律和统计理论有机地结合起来。

(4) 信息预测法

信息预测法是利用计算机技术组建大容量的数据库,包括各种农业措施、病虫害资料、气候条件等多种因素,然后利用系统分析技术、计算机辅助决策系统、数理统计手段和图形处理技术对众多因子进行最优设计处理。建立计算机管理系统对病虫害的预测,主要包括病虫害数据库管理系统、病虫害发生时间预测系统和信息传递系统、管理信息系统、决策支持系统、专家系统和地理信息系统。

(5) 组合预测法

组合预测法是设法把不同的预测模型组合起来,产生一个新的模型。它是在诸种单项预测模型各异且数据来源不同的情况下,将各种不同类型的单项模型兼收并蓄、各取所需,集中更多的信息与预测技巧,以减少预测的系统误差,显著改进预测效果。组合预测法的最主要特点是能最大限度地使用信息,以防止差错、消除虚假现象和避免一些不合理的假设。

1.4.2 病虫害预警

病虫害的发生、发展与气候因素、地理因素和植物品种之间有着密切的关系,它同时受到多种因素的影响,是一个复杂的非线性系统。影响病虫害发生的内部因子和外部环境之间存在着复杂的相互作用,许多复杂、不确定、模糊、随机的现象交织在一起,若单纯采用传统的数据处理方法,将很难解释并表述复杂病虫害系统的内部规律。

随着人工智能的发展,交叉学科不断涌现,同时出现了许多新的数学模型理论和优化技术,如人工神经网络技术、小波分析技术、混沌理论、支持向量机理论、智能优化算法,这些理论



为解决病虫害预警系统的复杂问题提供了有效的工具和手段,也使复杂繁琐的数据处理成为可能。[26,27]

1.5　研究的创新点及其成果意义

本书的研究领域涉及农业、水产、信息、管理等学科(专业),不仅能对这些学科发展起到很大的促进作用,而且还能有效地促进农业技术的推广。

1.5.1　研究的创新点

本书总结的科研工作成果有以下创新。

(1) 理论方法创新。提出了一套基于诊断理论、预警理论和会诊理论的能够指导农业病虫害远程诊断和预警的理论体系,为病虫害诊断与预警研究提供了理论基础。

(2) 技术集成创新。将人工智能技术和三层 B/S 技术、ASP、COM 组件技术、Windows DNA 技术、Java 技术、呼叫中心技术、移动 PDA 技术有效集成起来,实现了水产养殖中疾病诊断、牛病、棉花等农作物信息的全方位系统集成;还尝试性地将呼叫中心技术以及移动通信技术应用到网络化水产养殖专家系统的推广与应用过程中,扩大了系统的应用与推广智能化水产养殖专家系统平台的基础。

(3) 应用领域创新。完成了鱼病诊断专家系统、牛病诊断专家系统、棉花病害诊断专家系统、果树病害诊断系统等,涉及了农业病虫害诊断的多个领域;开创性地研究了基于 Web 的鱼病诊断系统、基于 PDA 的移动式鱼病诊断专家系统、基于呼叫中心的鱼病诊断专家系统以及鱼病远程会诊等应用系统,大大扩展了智能化水产养殖专家系统的平台,填补了相应领域的空白。

(4) 推广模式创新。针对我国农村网络技术普及的实际状况,开发了基于呼叫中心的鱼病远程咨询系统和移动式鱼病诊断专家系统,实现了网上诊断决策系统、远程会诊、呼叫中心和移动式诊断决策系统 4 种模式的鱼病诊断和防治体系,建立了一种 24 小时全天候的服务咨询模式,为水产养殖全过程管理提供了一种快捷、方便、直接的科技推广平台与管理模式。

1.5.2　对相应研发工作的影响

本书总结的科研工作成果对相应研发工作的影响表现在以下方面。

(1) 在农作物保护研究方面,病、虫、草、鼠害防治是维护农业生态系统健康的重要问题。该项目的研究促进了多层面地发展农作物医学信息技术迈开新的步伐,以辅助进行农业生态系统的健康普查、疾病诊断、风险评估和适合度判定。

(2) 在渔业疾病发生与防治研究方面,该研究可以有效促进现代信息技术、生物技术与渔业生态系统健康养殖理论的结合,开展水产养殖综合病害防治技术、无公害健康养殖技术、绿色生物鱼药及生物制剂开发技术等水产高新技术的研究和产业化开发,为进行健康养殖工程技术研究与成果转化提供了良好条件。

(3) 在 3S(GPS,GIS,RS)技术与相关信息技术的研究与应用方面,该研究可以促进形成基于 3S 技术进行病、虫、草害状态信息和农田生态环境信息的快速采集技术,结合空间分布信息快速处理理论与方法,研制开发基于 GIS 的病虫害综合治理系统,实现知识库、模型库、推理机与地理信息库的有机动态结合,动态显示病虫害的灾情、险情地理分布图,直观性强,并且

利于掌握病虫害的全局地域分布规律，从而进行有效防治，对作物病虫害的发生趋势作出准确预报，为农业部门和生产者提供科学的防治决策。

（4）在农业技术推广与服务模型研究方面，该研究可以有效地结合数字音视频、远程通信、数据库管理等信息技术，实现可视化、实时交互、及时应答的农业技术服务模式，有效、及时、合理地调度有限的专家资源，并实现跨地域的实时农业技术服务，提高农业技术服务的水平和能力。

（5）在提高农户知识能力方面，农业病虫害知识的推广和应用，提高了农户对病虫害发生、发展、预测、防治的防范意识和控制能力。特别是利用信息化手段对农户进行培训与教育，不仅较大程度地降低了病虫害发生的频率和危害的程度，促进了农业可持续发展，有利于农业生态环境保护，而且也提升了农户在信息时代的信息与知识管理能力。

1.5.3 对相应学科发展的影响

本书涉及的科研工作成果促进了医疗信息学的理念在农业病虫害诊断领域的实现，孕育了农业诊断信息学。在项目研究过程中，课题组成员运用计算机、通信、信息系统等信息技术以及统计分析、计量分析等定量分析手段，首次对农业病虫害诊断中的诊断流程，疾病、症状、病因各自的规律以及三者之间的相互关系进行了定量化模拟与分析，实现了运用计算机技术来处理农业病虫害诊断信息，包括诊断、教学和科研信息的收集、整理、传递、存储和检索，这些都为农业诊断信息学的产生提供了良好的研究基础。

本书涉及的研究成果促进了农业工程与信息工程的相互交叉，有效地加速了农业工程学科中精细农业方向的发展。目前，精细农业是近年来国内外在农业生产管理与农业实践中的热点研究课题，精细农业的核心技术之一是通信与决策支持系统。本课题的研究成果将会对目前精细农业研究与实践中的瓶颈问题——“提供合理的决策处方”的解决提供技术支持，促进精细农业中关键适用技术在我国的推广。

1.6 本书的研究目的和内容

1.6.1 研究目的

我们的研究目的是系统总结病虫害远程诊断和预警技术相关概念、理论、方法、技术和应用，并突出以下四方面的工作：①理论体系。提出一套基于诊断理论、预警理论和群决策理论基础的能够指导农业病虫害远程诊断和预警的理论体系，为病虫害诊断与预警研究提供理论方法。②方法模型。系统研究农业病虫害诊断和预警的知识表示、知识推理、知识获取以及群决策模型，探索获取农业病虫害诊断知识，准确、真实表达农业病虫害知识，病虫害预测、预警模型，有效求解农业病虫害诊断问题以及处理疑难杂症问题的解决模型，为相关人员提供方法支持。③技术支持。结合网络技术、计算机技术和数据分析技术、短信技术、移动 PDA 技术、视频会议技术、呼叫中心技术、预警技术与 Web 专家系统技术，系统总结国内外，特别是著者所在研究团队在农业病虫害诊断领域的应用探索经验，为相关人员提供技术支持。④示范和应用。在上述研究成果的基础上，结合软件工程实现农业病虫害诊断和预警专家系统的开发和示范应用。



1.6.2 研究内容

如同序言中所指出的,本书内容分为理论篇、方法篇、应用篇和示范篇四篇,内容框架如图 1-1 所示。

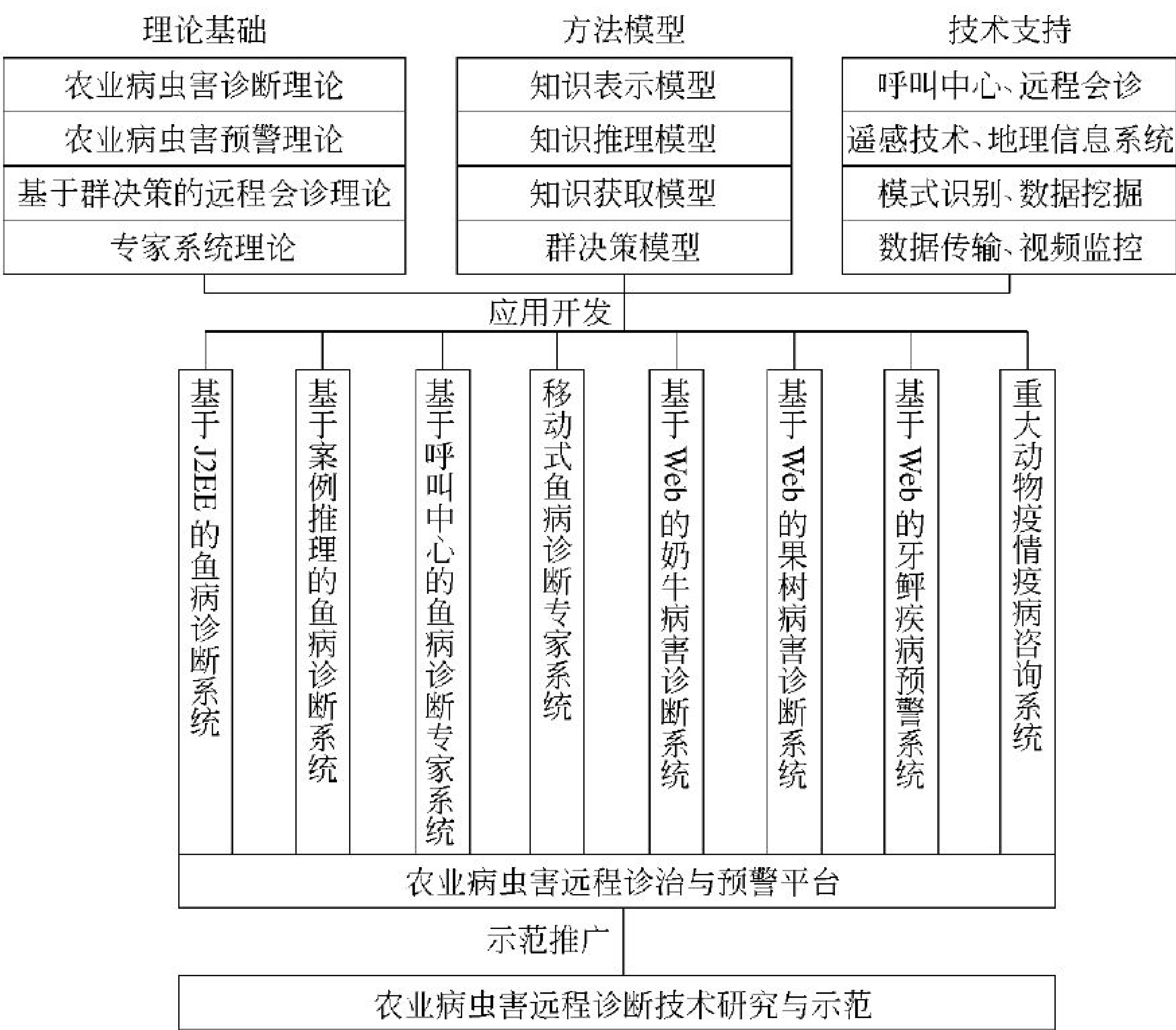


图 1-1 本书的内容框架

参考文献

1. 钟东屏. 诊断逻辑学[M]. 贵州:贵州科技出版社,1991
2. 丁洪. 基于物理过程描述的诊断推理[D]. 华中科技大学学报(自然科学版),1991,2
3. 徐越. 反绎推理和诊断问题求解的研究[D]. 中国科学院软件研究所,1995
4. 高济. 基于知识的软件智能化技术[M]. 杭州:浙江大学出版社,2000
5. 杨叔子,丁洪,史铁林等. 基于知识的诊断推理[M]. 北京:清华大学出版社,1993
6. 彭国强. 基于因果网络的诊断系统的研究[D]. 中国科学院软件研究所,1998
7. 温继文. 基于知识的鱼病诊断推理系统研究[D]. 中国农业大学,2003
8. 肖南峰,何彪等. 面向疾病预报、诊断、决策和监测的智能系统[J]. 华南理工大学学报,1997. 10
9. 杭小树等. 基于 CBR 的农作物病虫害预报专家系统[J]. 计算机工程与应用,2000. 10
10. 丁克坚等. 水稻主要病害诊断、预测、防治专家系统的研究[J]. 安徽农业大学学报,1998. 25(2):133~137
11. 刘洋. 基于多方法协作的玉米病虫害防治多媒体专家系统[D]. 哈尔滨工业大学,1997
12. 王亚东,陶海东等. 大豆病虫害诊断专家系统中知识库的建造[J]. 计算机与农业,2000(5),34~38
13. 杨盘洪,陈慧君. 基于 ASP 技术的小麦专家系统的研制. 太原理工大学学报,2000. 5,233~236
14. 柴毅,黄席樾,何离庆等. 番茄栽培病虫害防治知识表示. 重庆大学学报,2000. 11,56~58
15. 汪明磊,余为一等. 鸡病诊断专家系统的知识表示与推理机制的探讨. 计算机与农业,2001. 1,16~18

16. 郑育红，傅泽田等. 鱼病诊断专家系统设计[J]. 中国农业大学学报，2000. 5(6)：94～96
17. 杨萍著. 鱼病远程会诊群决策支持系统研究[D]. 中国农业大学，2006
18. DeSanctis G，Gallupe RB. A foundation for the study of group decision support systems [J]. Management Science，1987，33(5)：589～609
19. Dubois D，Prade H. Representation and combination of uncertainty with belief functions and possibility measures [J]. Computational Intelligence，1998，(4)：244～264
20. Bui TX(席酉民，侯宁霞译). 群体决策支持系统[M]. 西安：西安交通大学出版社，1992
21. 史美林，向勇，杨光信等. 计算机支持的协同工作理论与应用[M]. 北京：电子工业出版社，2000
22. 李军怀，周明全，耿国华. 远程医疗的国内外现状及展望[J]. 国外医学生物医学工程分册，2002，(5)：193～195，202
23. 蒋金根，郑辉. 远程医疗会诊 3350 例分析[J]. 中国新医药，2003，8：22～24
24. 马飞，程遐年. 害虫预测预报研究进展(综述). 安徽农业大学学报，2001，28(1)：92～97
25. 张孝羲，李运臂. 昆虫生态及预测预报[M]. 北京：农业出版社，1985，300～333
26. 王杰，檀根甲，胡易冰等. 基于神经网络的稻白叶枯病中期预警[J]. 安徽农业大学学报，2002，29(1)：12～15
27. 高宁. 基于 BP 神经网络的农作物虫情预测预报及其 MATLAB 实现[D]. 安徽农业大学，2003

第 2 章

农业病虫害诊断理论

要有效解决农业病虫害诊断领域的问题和实现诊断智能化，领域专家必须借助概念模型，把领域知识准确地传达给开发人员，而开发人员通过概念模型可以选择合理的知识表示和推理模型。为此，本章将现实世界的农业病虫害诊断问题、诊断知识和诊断求解方式进行抽象与提炼，建立农业病虫害诊断概念模型。本章的研究为后续建立农业病虫害诊断专家系统提供了理论基础。

2.1 农业病虫害诊断概念模型

概念模型是对真实世界的抽象描述，严格、完备的概念模型在领域专家与开发人员的知识沟通过程中具有不可替代的作用。[1]在农业病虫害诊断领域，专家借助概念模型把自己的领域知识准确地传达给开发人员，而开发人员通过概念模型可以选择合理的知识表示和推理模型，为构造合理的农业病虫害诊断专家系统提供重要的理论基础。所以，将现实世界的农业病虫害诊断问题进行识别和定义，通过对病虫害诊断问题、诊断知识和诊断求解方式进行抽象与提炼，建立病虫害诊断概念模型，是构造合理的诊断专家系统的重要基础。

本章首先对诊断和诊断问题进行统一定义，为农业病虫害诊断推理的研究提供理论依据。

2.1.1 诊断与诊断问题的定义

诊断是指诊断者对诊断对象，从其症状表现出发，根据诊断知识给出诊断结论的过程，是一个思维推理过程。[2]尽管诊断问题因领域不同而呈现出多样性，但是人们解决诊断问题的方式和方法却有极大的相似性。首先对诊断和诊断问题进行统一定义，然后在此基础上来研究农业病虫害诊断问题的概念体系。

诊断推理分别涉及诊断对象、故障、征兆、诊断知识和诊断结论，[3]下面一一阐述。

(1) 诊断对象：指诊断所作用的对象，由于诊断对象往往可以视为一个系统，所以又称作诊断系统。诊断系统通常由诊断者根据需要加以界定，诊断系统的界定指明了诊断所需要的知识范围和诊断思维可能延伸的范围。

(2) 故障：诊断系统正常与否，是由其基本组成部分正常与否决定的，系统正常的充要条件是每个基本组成部分都是正常的。从理论上讲，每个基本组成部分都有可能发生异常，把发生异常的基本组成部分称为基本故障。

(3) 征兆：诊断系统正常与否，通常要从一些确定的方面考察，每个方面都有正常与不正常之分。征兆表现是诊断系统发生故障的标志。

(4) 诊断表现：已知的征兆表现，它是诊断的依据，离开诊断表现，诊断无从说起。

(5) 诊断知识：包括系统描述知识(系统的结构知识、行为知识及功能知识)和系统关系知识(包括故障集 Faults=$\{f_i | i=1,2,\cdots,n\}$;症状集 Manifestation=$\{m_i | i=1,2,\cdots,n\}$以及它们的关系知识 $R \subseteq$ Faults×Manifestation);系统观测知识(系统已知的症状知识和系统输出的诊断知识)。

(6) 诊断结论：指对诊断系统发生的故障所作出的推断。对于一组表现,给出的诊断结论应该满足下述条件：在肯定诊断结论的情况下,诊断系统的表现可以得到最好解释,即在诊断结论成立的条件下,诊断系统的症状表现与诊断结论无矛盾,具有一致性。

(7) 诊断推理：指诊断者对所给诊断问题,从系统表现的症状出发,根据诊断知识,给出诊断对象诊断结论的过程,是一个思维推理过程。在这个过程中,采用什么样的推理是由诊断知识的形成特点决定的。如果诊断知识用于描述故障与特征关系的部分是从特征到故障,那么诊断推理的主要形式是演绎推理。反之,若诊断知识中用于描述故障与特征关系的部分是从故障到特征,那么诊断推理的主要形式是反绎推理。

(8) 一个诊断问题及其求解可以这样表示：

$$\mathrm{DC} = f(\mathrm{DP}) = f(\mathrm{DO},\mathrm{DF},\mathrm{DM},\mathrm{DK},\mathrm{DD})$$

其中符号的含义如下：

诊断问题——DP(diagnose problem);

诊断对象——DO(diagnose object);

故障集——DF(diagnose faults)=$\{d_1,d_2,\cdots,d_m\}$;

征兆集——DM(diagnose manifestation)=$\{m_1,m_2,\cdots,m_n\}$;

诊断知识——DK(diagnose knowledge);

诊断表现——DD(diagnose display)=DM^+,$\mathrm{DM}^+ \subseteq \mathrm{DM}$;

诊断结论——DC(diagnose conclusion),$\mathrm{DC} \subseteq \mathrm{DF}$;

诊断推理——f。

诊断问题的求解目标是给出诊断结论 DC,诊断的依据是 DK 和 DD,诊断过程 f 即推理过程。这里给出的诊断问题是一个一般性定义,除 DD,DM,DM^+ 确定之外,DO 与 DK 没有给出确定的形式。随着研究对象的不同,DO 与 DK 的表示方法会随之变化,从而可出现多种形式的诊断问题,导致多种模型的出现。

(9) 诊断问题的求解方式：是假设驱动的“假设—验证”循环过程,在每次循环中都包含如下两步,即①假设：寻找当前观测表现下可能的诊断假设;②验证：进行新测量,获取更多信息,对①中的假设进行验证。

2.1.2　农业病虫害诊断概念模型

在上述诊断及诊断问题定义的基础上，将现实世界农业病虫害诊断问题的主要因素抽象出来，明确各因素之间的关系，建立相应的农业病虫害诊断概念模型。

农业病虫害诊断是指农业病虫害专家根据诊断对象所表现出的特征信息,采用一定的诊断方法对其进行识别,以判定客体是否处于健康状态,找出相应原因并提出改变状态或预防发生的办法,从而对客体状态作出合乎客观实际的结论的过程。[4]

定义 2.1(关于农业病虫害诊断对象的定义)　农业病虫害诊断对象是由生活在大自然环境中的植物(或动物)构成的生态系统,主要包括畜禽、水产、农作物、果树等。

植物(或动物)直接或间接受到周围环境的影响,另外还受自然环境、人为因素的影响。因



此，在诊断中要重视环境对植物体（或动物体）本身的影响，避免割断联系的静态认识与分析方法的缺陷。

定义2.2（关于病因的定义）　病因是导致农业病虫害诊断对象系统处于不正常状态的原因。

某种植物（或动物）系统的致病因素复杂繁多，既有内因又有外因；既有直接致病因素，又有间接诱发因素。

令 $C=\{c_1,c_2,\cdots,c_k\}$ 代表有限非空的农业病虫害病因集合，$c_i(i=1,2,\cdots,k)$ 代表某种病因。

定义2.3（关于农业病虫害的定义）　农业病虫害是病体与外界生存环境致病因素相互作用，致使平衡破坏的结果。每种农业病虫害都是对生态系统中体现在病体上的一组特定的、不正常的状态的描述。

令 $D=\{d_1,d_2,\cdots,d_m\}$ 代表有限非空的农业病虫害集合，$d_i(i=1,2,\cdots,m)$ 代表某种农业病虫害。

定义2.4（关于症状的定义）　症状是指病体在发病过程中，出现的背离正常生理范围的、可以直接用于疾病诊断的信息。

令 $M=\{m_1,m_2,\cdots,m_n\}$ 代表有限非空的农业病虫害系统症状集合，$m_i(i=1,2,\cdots,n)$ 代表某种症状。

定义2.5（关于农业病虫害诊断知识的定义）　农业病虫害诊断知识是指症状、疾病、病因及其三者之间的关系的知识。

令 $K=\{C\cup D\cup M\}$ 代表农业病虫害诊断对象中病因、疾病、症状三者及其相互之间的关系知识。

定义2.6（关于农业病虫害诊断结论的定义）　专家通过对农业病虫害诊断对象特征信息的识别，找出其处于不正常状态的原因，并提出使其恢复正常状态的方法。

通过上述农业病虫害诊断中一些概念的定义，给出农业病虫害诊断问题的如下定义。

定义2.7（关于农业病虫害诊断问题的定义）　农业病虫害诊断问题可以采用下列表达形式：

$$\mathrm{DC}=f(\mathrm{DP})=f(\mathrm{DO},D,C,M,K,M^{+})$$

其中符号的含义如下：

诊断问题——DP，这里指农业病虫害诊断问题；

诊断对象——DO，这里指生活在自然环境中的某种受害植物（或动物）；

疾病集—— $D=\{d_1,d_2,\cdots,d_m\}$ 指某种植物（或动物）所有可能发生的疾病集合；

症状集—— $M=\{m_1,m_2,\cdots,m_n\}$ 指某种植物（或动物）所有可能表现的症状集合；

病因集——$C=\{c_1,c_2,\cdots,c_k\}$ 表示诊断系统中所有病因的非空有限集合；

诊断知识——$K=\{C\cup D\cup M\}$ 指病因、疾病、症状及其三者之间关系的知识；

诊断表现——$M^{+}\subseteq M$ 表示观测到的症状集合；

诊断结论——DC，$\mathrm{DC}\subseteq C$；

诊断推理——f，指农业病虫害诊断专家的思维推理过程。

农业病虫害诊断问题的实质是专家通过症状集 $M^{+}\subseteq M$，依据诊断知识 $K=\{C\cup D\cup M\}$，运用诊断推理 f，求解得出结论 DC 的过程。

2.2 农业病虫害诊断的数学定义

农业病虫害的传播是初始病因侵袭健康机体引起原发性疾病，原发性疾病成为病源引起继发性疾病，继发性疾病成为新的疾病源，引起另一疾病的发生，从而导致疾病的原因与结果相互转化，形成互为因果、不断转化的因果网络的过程。[5]因此，可有以下定义。

定义 2.8 农业病虫害诊断系统的诊断问题可以用一个五元式 $DP=\langle D,M,C,A,M^*\rangle$ 来描述，其中：

$D=\{D_1,D_2,\cdots,D_L\}$ 表示系统中各层次系统的所有疾病的非空有限集合；

$M=\{M_1,M_2,\cdots,M_L\}$ 表示系统中各层次系统的所有病症的非空有限集合；

$C=\{C_1,C_2,\cdots,C_L\}\subseteq D\times M$ 表示 D 到 M 之间的因果映射关系集合；

$A=\{A_1,A_2,\cdots,A_L\}\subseteq M\times D$ 表示 M 到 D 之间的因果映射关系集合；

$M^*\subseteq M$ 表示已知的观测病症集合。

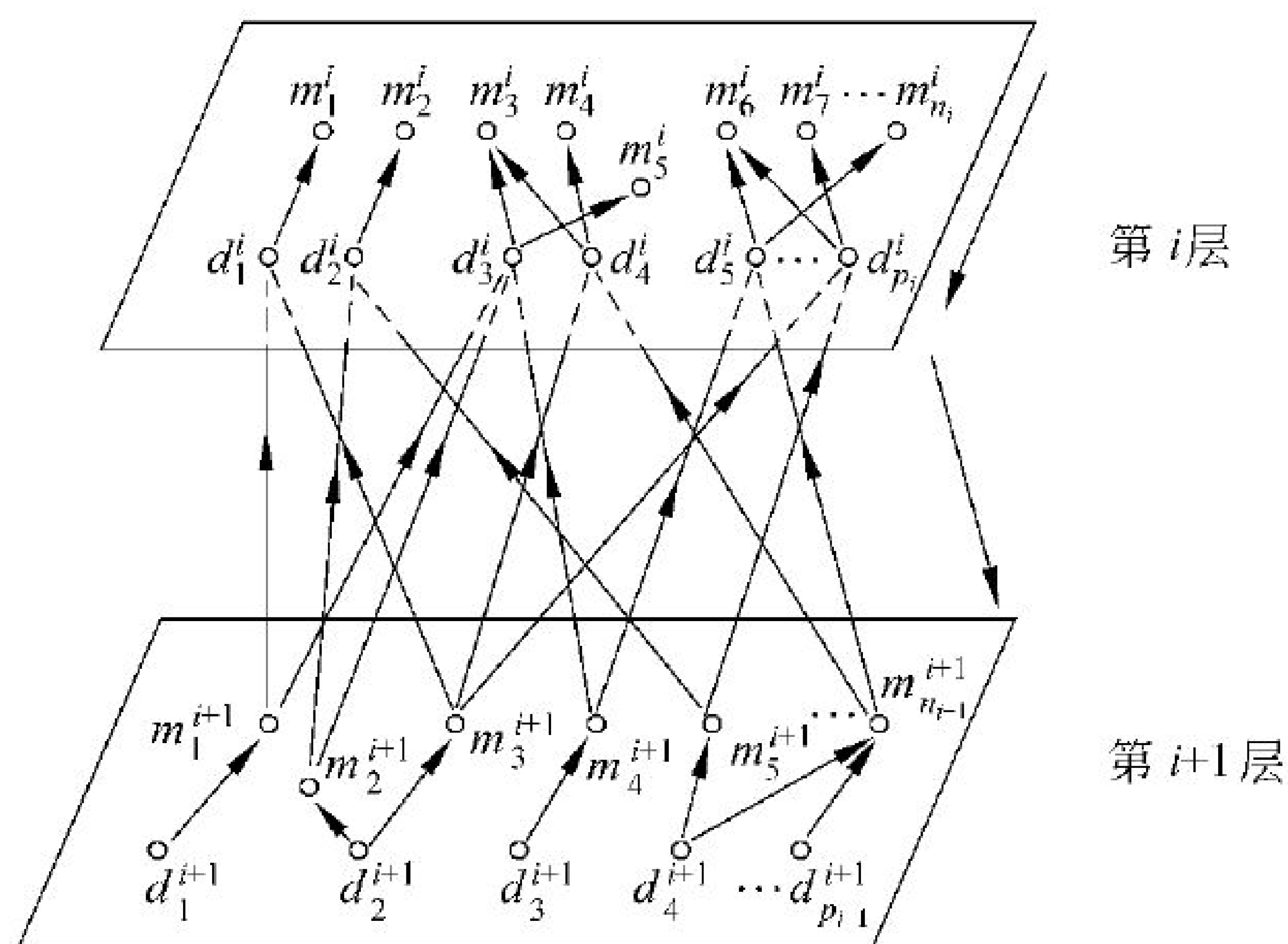


图 2-1 “病因→疾病”和“疾病→病症”两种因果映射关系

由定义 2.8 可知，D_i 表示第 i 层次的疾病集合；M_i 表示第 i 层次的病症集合；L 表示复杂系统的层次深度；关系 C_i 刻画了各层次中由疾病到病症的因果映射关系，即 $C_i=\{\langle d_k^i,m_j^i\rangle \mid d_k^i\in D_i,m_j^i\in M_j,i=1,2,\cdots,L,k=1,2,\cdots,p_i,j=1,2,\cdots,n_i\}$，其中 $\langle d_k^i,m_j^i\rangle$ 表示第 i 层次系统中第 k 个疾病 d_k^i 的产生可能会引起同一层次的第 j 个病症 m_j^i 关系，A_i 则描述了低一层次系统的病症集合 M_{i+1}（作为病因）到上一层次系统的疾病集合 D_i 之间的因果映射关系，即 $A_i=\{\langle m_j^{i+1},d_k^i\rangle \mid m_j^{i+1}\in M_{i+1},d_k^i\in D_i,i=1,2,\cdots,L,k=1,2,\cdots,p_i,j=1,2,\cdots,n_i\}$，其中 $\langle m_j^{i+1},d_k^i\rangle$ 表示第 $i+1$ 层次系统的第 j 个病症 m_j^{i+1} 可能是第 i 层次系统的第 k 个疾病 d_k^i 产生的病因；$M^*=\left(\bigwedge_{m_j\in M} m_j\right)\subseteq M$ 为已知的系统级和其他层次级的观测病症集合。图 2-1 描述了上述两类因果映射关系。图中的层次关系表示了变量间的因果依赖关系：上层依赖于下层，图中不相关变量组（不在同一框内的变量组）的状态转移可以组合形成多个新的状态。在层次结构定义中，同一层次由许多变量组组成，变量组内的变量为相关变量，它们的状态变化直接影响其上层变量，而不同变量组内的变量可以在不影响各自上层变量的前提下相互组合。

根据上述两类映射关系，给出如下定义疾病、病症和病因三种映射函数的定义 2.9。

定义 2.9 对于任一 $m_j^i\in M_i,m_j^{i+1}\in M_{i+1},i=1,2,\cdots,L-1,k=1,2,\cdots,p_i,j=1,2,\cdots,$



$n_i\}$，定义两种情况下的疾病映射函数为

$$D(m_j^i)=\{d_k^i \mid \langle d_k^i, m_j^i\rangle \in C_j\}$$
$$D(m_j^{i+1})=\{d_k^i \mid \langle m_k^{i+1}, d_k^i\rangle \in A_i\}$$

定义病症和病因映射函数分别为

$$M(d_k^i)=\{m_j^i \mid \langle d_k^i, m_j^i\rangle \in C_i\}$$
$$R(d_k^i)=\{m_l^{i+1} \mid \langle m_l^{i+1}, d_k^i\rangle \in A_i\},\quad l=1,2,\cdots,n_{i+1}$$

由定义 2.9 可以看出，在复杂系统诊断中，第$(i+1)$层次的所有病症集合即为第 i 层次的所有疾病的病因集合，直观上，这是较为合理的。相应地，给出如下基本事件的定义 2.10。

定义 2.10　对于系统诊断问题 $\mathrm{DP}=\langle D,M,C,A,M^*\rangle$：

① 疾病事件　若疾病 $d_k^i\in D_i(i=1,2,\cdots,L;k=1,2,\cdots,p_i)$，则表示疾病 d_k^i 存在，而同一层次的其他疾病是否存在未知。类似地，$\overline{d_k^i}$仅表示 d_k^i 不存在，而同一层次的其他疾病是否也不存在则未知。

② 征兆事件　若征兆 $m_j^i\in M_i(i=1,2,\cdots,L;j=1,2,\cdots,n_i)$，则表示征兆 m_j^i 存在，而同一层次的其他征兆是否存在未知。类似地，$\overline{m_j^i}\in\overline{M_i}$仅表示 m_j^i 不存在，而同一层次的其他征兆是否也不存在则未知。

③ 因果事件　对于任何 $d_k^i\in D_i, m_j^i\in M_i, m_j^{i+1}\in M_{i+1}(i=1,2,\cdots,L;j=1,2,\cdots,n_i;k=1,2,\cdots,p_i)$，符号 $m_j^i:d_k^i$ 表示疾病 d_k^i 确实引起了征兆 m_j^i；而符号 $d_k^i:m_l^{i+1}$ 表示征兆 m_l^{i+1} 作为原因确实导致疾病 d_k^i 的产生。显然，只有当 m_j^i, d_k^i 同时发生，并且 d_k^i 确实引起 m_j^i 时，因果事件 $m_j^i:d_k^i$ 才是真的(存在)；类似地，只有当 d_k^i, m_l^{i+1} 同时发生，并且 m_l^{i+1} 确实导致 d_k^i 时，因果事件 $d_{k+1}^i:m_l^{i+1}$ 才存在。

该定义主要表现在层次性、征兆与原因的统一和两种类型的因果映射关系的统一方面。可知定义 2.10 更符合复杂系统诊断问题的实际情况。

定义 2.11　因果事件$\langle m_j^i:d_k^i\rangle(i=1,2,\cdots,L;j=1,2,\cdots,n_i;k=1,2,\cdots,p_i)$的一个环境定义为 D_i 和 C_i 中除了$\langle m_j^i:d_k^i\rangle$及其否定事件$\overline{\langle m_j^i:d_k^i\rangle}$以外的其他任何疾病事件和因果事件的合取事件；类似地，因果事件$\langle d_k^i:m_l^{i+1}\rangle$的一个环境被定义为 M_{i+1} 和 A_i 中除了$\langle d_k^i:m_l^{i+1}\rangle(l=1,2,\cdots,n_{i+1})$及其否定事件$\overline{\langle d_k^i:m_l^{i+1}\rangle}(l=1,2,\cdots,n_{i+1})$以外的其他任何病症事件和因果事件的合取事件。

在此基础上，可正式给出层次因果模型所要求的如下有关假设。

假设 2.1　概率知识假设。对于复杂系统中最深层的疾病事件 $d_k^L\in D_L(k=1,2,\cdots,p_i)$，它们的先验概率 $P(d_k^L)$已知，且 $0\leqslant P(d_k^L)<1$；对于所有的因果事件$\langle m_j^i:d_k^i\rangle$，其条件因果概率 $P(\langle m_j^i:d_k^i\rangle|d_k^i)$已知，且满足 $0\leqslant P(\langle m_j^i:d_k^i\rangle|d_k^i)\leqslant 1(i=1,2,\cdots,L;k=1,2,\cdots,p_i;j=1,2,\cdots,n_i)$，如果$\langle d_k^i,m_j^i\rangle\in C_i$，则 $P(\langle m_j^i:d_k^i\rangle|d_k^i)>0$；类似地，对于所有的因果事件$\langle d_k^i:m_l^{i+1}\rangle(l=1,2,\cdots,n_{i+1})$，其条件因果概率 $P(\langle d_k^i:m_l^{i+1}\rangle|m_l^{i+1})$已知，且满足 $0\leqslant P(\langle d_k^i:m_l^{i+1}\rangle|m_l^{i+1})\leqslant 1, i=1,2,\cdots,L-1$，如果$\langle m_l^{i+1}:d_k^i\rangle\in A_i$，则 $P(\langle d_k^i:m_l^{i+1}\rangle|m_l^{i+1})>0$。

假设 2.2　最深层疾病事件独立假设。对于复杂系统中最深层的所有疾病事件 D_L，假设一个疾病事件 $d_k^L\in D_L, k=1,2,\cdots,p_L$ 独立于其他疾病事件 $d_q^L\in D_L(q=1,2,\cdots,p_L)$而发生，即存在

$$P\Big(\bigwedge_{d_k^L\in D_L} d_k^L\Big)\in\prod_{d_k^L\in D_L}P(d_k^L)$$

在以往的多疾病诊断问题求解中,几乎都假设所有疾病事件的发生是相互独立的。这显然不尽合理,因为复杂系统中很多高层次子系统的疾病事件的发生是非独立的。这里只假设复杂系统中最深层系统中的所有疾病事件相互独立,如此更加逼近疾病产生与传播的实际情况。

假设 2.3 同层因果事件独立假设。对于任一 $d_q^i \in D_i, i=1,2,\cdots,L, q=1,2,\cdots,p_i$,因果事件$\langle m_j^i : d_k^i \rangle, j=1,2,\cdots,n_i, k=1,2,\cdots,p_i$ 的发生独立于其环境中任一事件的发生,假设 X 为$\langle m_j^i : d_k^i \rangle$的一个环境,且 $P(X \wedge d_k^i) \neq 0$,则

$$P(\langle m_j^i : d_k^i \rangle \mid d_k^i \wedge X) = P(\langle m_j^i : d_k^i \rangle \mid d_k^i)$$

类似地,假设存在

$$P(\langle d_k^i : m_l^{i+1} \rangle \mid m_l^{i+1} \wedge Y) = P(\langle d_k^i : m_l^{i+1} \rangle \mid m_l^{i+1})$$

其中 Y 为$\langle d_k^i : m_l^{i+1} \rangle$的一个环境,且 $P(Y \wedge m_l^{i+1}) \neq 0$。

假设 2.4 病症事件因果假设。对于所有层次的任何一个病症事件 $m_j^i \in M_j, i=1,2,\cdots, L-1, j=1,2,\cdots,n_i$,它的发生都是由相同层次的某些疾病事件引起的,即

$$m_j^i \leftrightarrow \bigvee_{d_k^i \in D(m_j^i)} \langle m_j^i : d_k^i \rangle, \quad i=1,2,\cdots,L-1, j=1,2,\cdots,n_i, k=1,2,\cdots,p_i$$

假设 2.5 疾病事件因果假设。对于最深层以外的其他所有层次的任何一个疾病事件 $d_k^i \in D_i, i=1,2,\cdots,L-1, k=1,2,\cdots,p_i$,它的发生都是由其下一层次的某些病症事件作为病因而导致的,即

$$d_k^i \leftrightarrow \bigvee_{d_k^{i+1} \in R(d_k^i)} \langle d_k^i : m_l^{i+1} \rangle, \quad i=1,2,\cdots,L-1$$

$$l=1,2,\cdots,n_{i+1}; k=1,2,\cdots,p_i; j=1,2,\cdots,n_i$$

2.3 农业病虫害诊断问题的求解模型

根据前面的基本定义和假设条件,针对农业病虫害诊断系统的实际情况,诊断问题求解过程可以分解为预处理、“疾病—病症”的层内诊断和“病因—疾病”的层间诊断三个子过程。预处理主要是将领域专家的实际经验知识作为约束条件,缩小问题求解的状态空间;层内诊断是指由已知的病症集合到产生该病症集合的该层中各子系统疾病集合的识别;而层间诊断则是由已知上一层的疾病集合,识别引起该疾病集合的病因集合,即输入到上一层子系统级的该层中各子系统的病症集合。[5]这样,复杂系统的诊断过程就成为这两类诊断子问题的循环求解过程,并以层内诊断作为该过程的开始和结束。

2.3.1 诊断知识的预处理

通过对专家长期积累的知识、经验及有关方面的科学资料进行分析,从而减少推理分支,避免组合爆炸。为了适应定性模型中的层次因果关系约束和搜索的实现,提高算法效率,对约束传播机制可以进行适当修改,[6]主要修改有:

(1) 增加状态产生的层次信息,即状态的代次信息。例如,初始化的所有状态是第一代状态,它们产生的后继状态是第二代,等等。

增加层次信息有两个目的:一是为了序列因果约束关系的实现;二是为了实现深度优先算法,以便基于目标搜索的实现。



（2）部分约束参与参数的状态转换，这样可大大减少变量的状态空间，使得有关过滤在较小的状态空间内实现，可以提高约束滤波效率。

（3）在约束过滤中增加序列因果关系约束过滤和目标状态的判断。

（4）采用递归算法，实现目标搜索。

2.3.2　层内诊断问题

第 i 层的层内诊断问题可用一个四元式 $\mathrm{DSPBH}_i=\langle D_i,M_i,C_i,M_i^*\rangle, i=1,2,\cdots,L$ 来描述，其中，$D_i=\{d_1^i,d_2^i,\cdots,d_{n_i}^i\}, i=1,2,\cdots,L$ 表示第 i 层各子系统所有疾病的非空有限集合；$M_i=\{m_1^i,m_2^i,\cdots,m_{p_i}^i\}, i=1,2,\cdots,L$ 表示第 i 层中疾病集合 D_i 所引起的所有病症的非空有限集合；C_i 是描述 D_i 到 M_i 之间的因果关系集合，即 $C_i=\{\langle d_k^i,m_j^i\rangle \mid d_k^i\in D_i, m_j^i\in M_i\}, i=1,2,\cdots,L, j=1,2,\cdots,n_i, k=1,2,\cdots,p_i$，其中 $\langle d_k^i,m_j^i\rangle$ 表示疾病 d_k^i 的产生可能会引起病症 m_j^i，并且 m_j^i 可能是异常的，也可能是正常的。如果 m_j^i 是一个病症，则 $\langle d_k^i,m_j^i\rangle\in C_i$ 表示疾病 d_k^i 的产生在一定程度上也会阻止病症 m_j^i 变为异常。这样，如果已知或观测到病症 m_j^i 是异常的，则这一事实也将反过来在一定程度上否定疾病 d_k^i 的产生，这种否定关系其实是一种很有用的诊断知识。M_i^* 为属于 M_i 的一个已知病症集合，它可通过测试或通过其他中间假设和结论推理获得，用于层内诊断的这些知识集合可表达如下：

$$\left.\begin{aligned} &M(d_k^i)=\{m_j^i \mid \langle d_k^i,m_j^i\rangle\in C_i\}, \forall d_k^i\in D_i\\ &D(m_j^i)=\{d_k^i \mid \langle d_k^i,m_j^i\rangle\in C_i\}, \forall m_j^i\in M_i\\ &M(D_i)=\bigcup_{d_k^i\in D_i} M(d_k^i)\end{aligned}\right\}\quad i=1,2,\cdots,L; j=1,2,\cdots n_i; k=1,2,\cdots,p_i$$

其中，$D_i^L\subseteq D_i$ 是一假设的疾病子集；并且

$$\left.\begin{aligned} &D(M_i^L)=\bigcup_{m_j^i\in D_i} D(m_j^i)\\ &M^*(d_k^i)=M(d_k^i)\cap M_i^*\\ &M^*(D_i^*)=M(D_i^*)\cap M_i^*\end{aligned}\right\}\quad i=1,2,\cdots,L; j=1,2,\cdots,n_i; k=1,2,\cdots,p_i$$

其中如果 $M_i^*\subseteq M(D_i^*)$，则称 D_i^* 为 M_i^* 的一个覆盖，根据节约覆盖理论，该疾病子集 D_i^* 必能解释（或引起）所有这些已知的病症集合。

对于第 i 层的层内诊断问题，即 $\mathrm{DSPWH}_i=\langle D_i,M_i,C_i,M_i^*\rangle, i=1,2,\cdots,L$ 当且仅当①$M_i^*\subseteq M(D_i^*)$，即 D_i^* 是 M_i^* 的一个覆盖；并且②D_i^* 是所有的覆盖集合中最节约的一个时，则 $D_i^*\subseteq D_i$ 是 DSPWH_i 的一个解。

2.3.3　层间诊断问题

从第 i 层到第 $(i+1)$ 层的层间诊断问题也可以用一个四元式来描述，即 $\mathrm{DSPBH}_i=\langle R_i, D_i,A_i,D_i^*\rangle, i=1,2,\cdots,L-1$，其中 R_i 是引起上述疾病子集 D_i 产生的病因的非空有限集合，事实上，该病因集合也正是第 $(i+1)$ 层次的病症集合 M_{i+1}，即 $R_i=M_{i+1}=\{m_1^{i+1},m_2^{i+1},\cdots, m_{n_i}^{i+1}\}(i=1,2,\cdots,L-1)$；$D_i$ 的含义与上述层内诊断相同；A_i 则是描述 M_{i+1} 到 D_i 之间的因果关系集合，即 $A_{i+1}=\{\langle m_j^{i+1},d_k^i\rangle \mid m_j^{i+1}\in M_{i+1}, d_k^i\in D_i\}(i=1,2,\cdots,L-1; k=1,2,\cdots,p_i; j=1, 2,\cdots,n_{i+1})$，其中 $\langle m_j^{i+1},d_k^i\rangle\in A_i$ 表示 m_j^{i+1} 可能是 d_k^i 产生的病因，或 m_j^{i+1} 的存在可能支持 d_k^i 产生这一假设。当 m_j^{i+1} 为病症时，则因果关系 $\langle m_j^{i+1},d_k^i\rangle\in A_i$ 表示疾病 d_k^i 的产生可能支持病症 m_j^{i+1} 是正常的这一假设；D_i 是上述定义的第 i 层的层内诊断问题的解。这样，层间诊断问

题的知识集合可表述如下：

$$\left.\begin{aligned} &D(m_j^{i+1}) = \{d_k^i \mid \langle m_j^{i+1}, d_k^i \rangle \in A_i\}, \forall m_j^{i+1} \in M_{i+1} \\ &R(d_k^i) = \{m_j^{i+1} \mid \langle m_j^{i+1}, d_k^i \rangle \in A_i\}, \forall d_k^i \in D_i \\ &D(M_{i+1}^*) = \bigcup_{m_j^{i+1} \in M_{i+1}^*} D(m_j^{i+1}) \end{aligned}\right\} \begin{aligned} &i = 1,2,\cdots,L-1; j = 1,2,\cdots,n_{i+1}; \\ &k = 1,2,\cdots,p_i \end{aligned}$$

其中，$M_{i+1}^* \subseteq M_{i+1}$是一个假设的病症子集；并且

$$\left.\begin{aligned} &R(D_i^*) = \bigcup_{d_k^i \in D_i^*} R(d_k^i) \\ &D^*(m_j^{i+1}) = D(m_j^{i+1}) \cap D_i^* \\ &D^*(M_{i+1}^*) = D(M_{i+1}^*) \cap D_i^* \end{aligned}\right\} \quad i = 1,2,\cdots,L-1; j = 1,2,\cdots,n_{i+1}; k = 1,2,\cdots,p_i$$

其中如果 $D_i^* \subseteq D(M_{i+1}^*)$，则 M_{i+1}^*也被定义为 D_i^* 的一个覆盖。

对于从第 i 层到第$(i+1)$层次的层间诊断问题 $DSPBH_i = \langle R_i, D_i, A_i, D_i^* \rangle$，当且仅当①$D_i^* \subseteq D(_{i+1}^*)$，即 M_{i+1}^*是 D_i^* 的一个覆盖；并且②M_{i+1}^*是所有 D_i^* 的覆盖中最节约的一个时，则 $M_{i+1}^* \subseteq M_{i+1}$是 $DSPBH_i$ 的一个解。

由此可见，对于复杂系统的诊断问题 $DP = \langle D, M, C, A, M^* \rangle$，第 i 层中疾病的有限子集 D_i 作为该诊断问题的解可以通过第 i 层的层内诊断和第$(i+1)$层的层间诊断的循环交叠求解获得。

2.4 小结

本章通过对农业病虫害诊断领域知识的概念化、诊断问题的形式化、诊断推理过程的逻辑化描述，对农业病虫害诊断知识进行透视分析，为更好地实现对领域知识的获取、表示、组织和问题求解以及系统的实现提供了理论基础。

参考文献

1. 温继文，李道亮，傅泽田等. 鱼病诊断专家系统中概念模型的构建[J]. 农业系统科学与综合研究，2006(8)：208～211
2. 钟东屏. 诊断逻辑学[M]. 贵州：贵州科技出版社，1991
3. 杨叔子，丁洪，史铁林等. 基于知识的诊断推理[M]. 北京：清华大学出版社，1993
4. 彭国强. 基于因果网络的诊断系统的研究[D]. 中国科学院软件研究所，1998
5. 温继文. 基于知识的鱼病诊断推理系统研究[D]. 中国农业大学，2003
6. 郭永洪. 基于本体的鱼病知识获取与诊断推理集成系统研究[D]. 中国农业大学，2004

第3章

农业病虫害预警理论

农业病虫害预警是依据生物学、生态学、数学原理，分析病虫害过去和现在的各种有关因素，判断病虫害的未来变化和发展趋势，以降低病虫害带来的灾害性损失。准确、及时的病虫害预警是有效进行病虫害诊疗和改善生态系统环境的基础。本章在分析研究农业病虫害预警预测内涵、要素和方法的基础上，阐述了病虫害预警指标体系和预警模型的构造过程。

3.1 农业病虫害预警

3.1.1 病虫害预警的概念与范畴

病虫害预警是指通过对环境、病原、个体本身等进行监测，分析所获得的监测数据，并根据影响因子之间的相互关系对未来个体的发病情况进行预测，如果有发病的可能，也就是预测的数据超过了正常的警限就进行预警。[1]

按预警范畴分类，农业病虫害预警主要分为农作物、畜禽、水产预警等（见图3-1），每一类农业病虫害预警因子都有所不同。[2]例如，农作物的主要预警因子包括：气温、湿度、光照等，而畜禽的主要预警因子包括传染源、空气的流通、畜禽交易、免疫程度等。

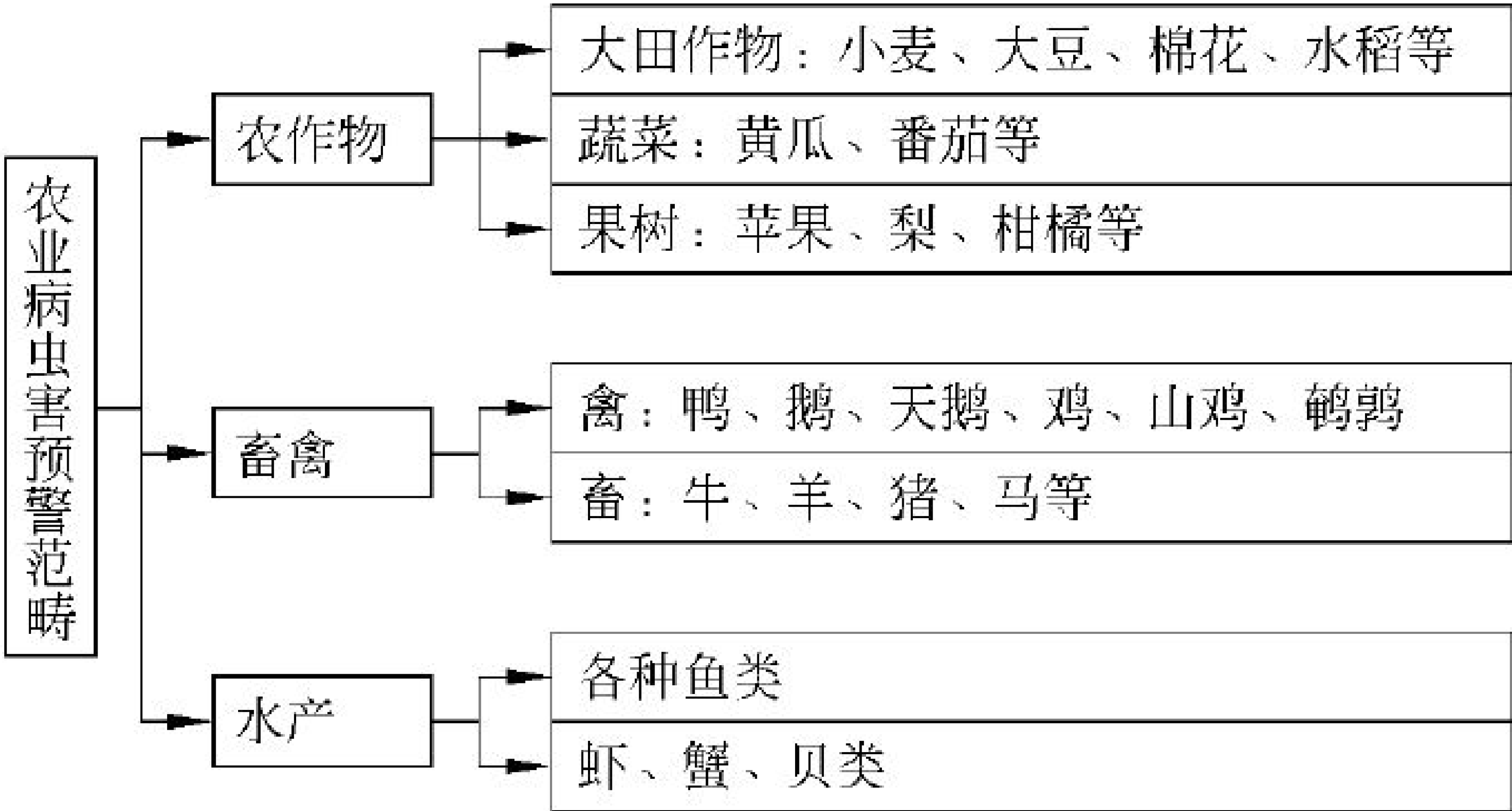


图3-1 农业病虫害预警范畴

3.1.2 病虫害的预警过程

农业病虫害的预警过程从逻辑上划分为确定警情、寻找警源、分析警兆、预报警度以及排除警情等一系列相互衔接的过程。确定警情是大前提，是预警研究的基础，而寻找警源和分析警兆属于对警情的因素分析及定量分析，预报警度则是预警目标所在，排除警情是目标实现的过程。[3]

（1）确定警情

这是农业病虫害预警过程的第一步。警情，就是事物发展过程中出现的异常情况，可以从两个方面考察，一是警素，指构成警情的指标。病害预警一般采用发生期、普遍率、严重度或病情指数作为警素。二是警度，即警情的程度。通常把警情的严重程度划分为五个警限，即无警、轻警、中警、重警、巨警。在这五种警限中，无警警限（又称安全警限）的确定是最为关键的。当警情指标的实际值不在安全警限范围内，则表明警情出现。结合具体情况，根据警情指标的实际值，观测其落在哪一警限区域，便可检测其警度。

（2）寻找警源

警源是警情产生的根源。对于植物病害流行系统来说，染病的寄主植物、具有致病性的病原物和有利于发病的环境、人类干预，即病害四面体，就是警情产生的根源。根据具体情况分析这四种警源分别是什么，对于每一种警源还可以再进一步细分，哪一种警源应作为研究的重点内容，需要具体问题具体分析。

（3）分析警兆

警兆是警情暴发的先兆，分析警兆是预警过程中的关键环节。从警源的产生到警情的暴发，其间必有警兆的出现。通常，不同的警情对应着不同的警兆。警兆可以是警源的扩散，也可以是警源扩散过程中其他相关的共生现象。通常，同一警情指标往往对应多个警兆指标，而同一警兆指标可能对应多个警情指标。当警情指标发生异常变化之前，总有一定的先兆（即警兆），这种先兆与警源可以有直接关系，也可以有间接关系；可以有明显关系，也可以有隐形的未知黑色关系。警兆的确定可以从警源入手，也可以依经验分析，分析警兆及其报警区间便可预报预测警情。确定警兆之后，需要进一步分析警兆与警情的数量关系，找出与警情的五种警限相对应的警区，警兆警区一般也分为五种情况：无警警区、轻警警区、中警警区、重警警区、巨警警区，然后借助于警区进行警度预报。

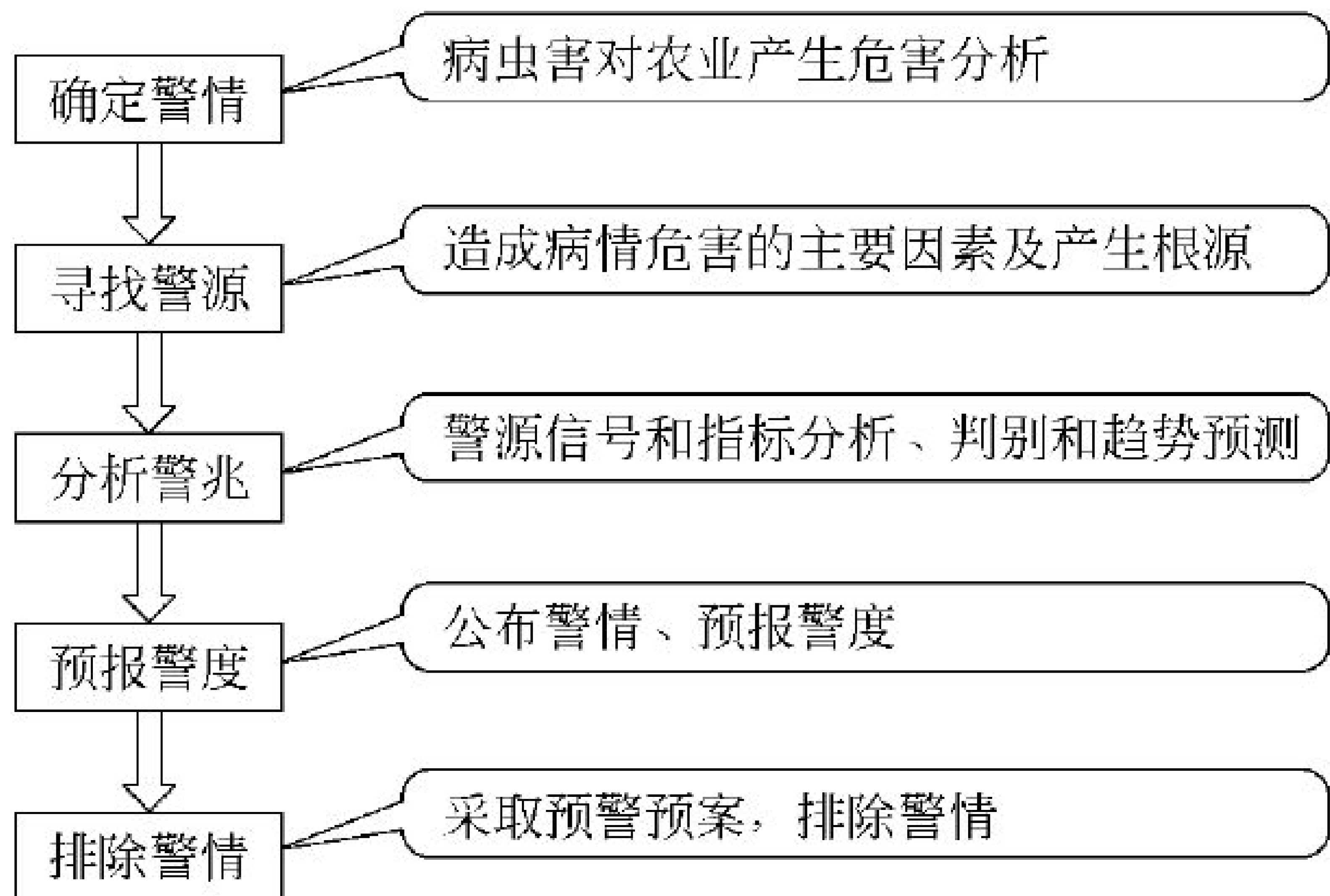


图 3-2　病害预警过程

（4）预报警度

预报警度是预警的目的。在预报警度中，需要注意结合经验方法、专家方法等，这样可提高报警精度。根据警兆的变动情况，结合警兆的变动区间，参照警情的警限或警情等级，运用定性和定量方法分析警兆报警区间与警情警限的实际关系，结合专家意见及经验，便可预报实际警情的严重程度。

（5）排除警情

针对每一种警情均给出相应的对策建议，以消除警情。对于病害预警来说，就是按照“预



防为主，综合防治，保护环境”的原则，向用户提供防治建议。

3.1.3 病虫害预警的特点

病虫害预警有以下特点。

（1）警情的累积性和突发性。农业病虫害是自然过程和人类活动影响下的结果，因此病虫害的发展、恶化程度会在时间、空间上有一个演替过程。在时间上，表现为从量变到质变的过程，包括渐变、突变、连续、间断、波动、周期、累积等各种演变形式；在空间上，则是系统内部各要素的消长、进退和更换、进化和退化。警情的累积性要求在农业病虫害预警分析中能涵盖一定的时间和空间范围；而警情的突发性则要求重在警情的预报，发现警兆并提出切实可行的措施。

（2）警兆的滞后性。由于警情的累积性特征，病虫害产生的后果显露要相对滞后一段时间，当警兆表现出来以后，警情的危害性已经相当深了。因此，需要加强对农产品及其环境的动态监测，实时监测被观测对象的状态，这样可以有效地预防病害。

（3）预警的动态性。评价的取值一般是静态的、一次性的，结论亦是一次性的，而预警的取值是多维的，是随不同时间而变化的序列，侧重对不同时间、时段动态变化的分析，可加强动态描述。预警评价可以把握不同时期内病害所处状态及其演化趋势和速度。

（4）预警的深刻性。预警的实现需要有评价和一般预测等大量工作作为基础，只有认识把握现状和演化趋势才能实现预警。预警阐明的环境问题对揭示环境本质及变化规律更为深刻、准确。可以说，预警研究的目的性、针对性更为集中、强烈，对环境的监督、管理作用更大，从而实现其警告与警示作用。

3.1.4 病虫害的预警方法

预警方法是指系统如何收集资料，进而如何分析整理资料，并提出预警信息，以至实现预警功能全过程的逻辑思想方法。目前有记载的预警方法有200种，按照不同的分类方式可有不同的预警方法。[2]

（1）从方法和手段角度可将预警方法分为指数法、专家经验法、模型法等

- 指数法：利用统计学的指数方法把警兆指标的变化综合成指数，依据一个或多个警兆指标指数的大小进行预警。警情指数公式为：$DP=\sum_{i=1}^{n}W_i$，其中DP表示警情指数，W_i表示第i个警兆指标的影响指数。
- 专家经验法：主要是根据过去的经验，采用仿真试验、主观衡量法、保守估计法、乐观悲观法、竞赛理论等数学模型用类比的方法推算出病虫害的变化趋势。是病虫害预警的初级方法，对病虫害发生情况的预警结果往往依赖于一些常见现象，无法精确定量地对病虫害发生情况进行预警。它是统计预警的基础，但病虫害与外界因子不是简单的线性关系，而经验法中所用的简单相关系数相关程度不高，做出的预警很难达到较高的准确率，同时由于预警时限短，不能满足人们制定长远规划的目的，因此经验预警法存在一定的局限性。
- 模型法：从狭义上讲是以警情为因变量，以警兆为自变量建立的预警模型；从广义上讲是预警系统中用以加工处理原始数据，形成并表达预警信息的方法。有数理统计法、系统动力学法、灰色系统法、人工神经网络法等。

(2) 按预警内容可将预警分为发生期预警、发生量预警和灾害程度预警

- 发生期预警：预警某种病虫害的状态或级别的出现期或危害期；对于具有迁飞、扩散习性的害虫，预警其迁出或迁入本地的时期，并以此作为确定防治适期的依据。方法包括历期法、期距法、物候法和有效积温法。
  - 历期法：根据某代或某虫态的发育历期，结合下段时间的温度等有关环境条件，推算下一代或下一虫态的发生期，以确定未来的防治适期、次数和防治方法。
  - 期距法：根据害虫某虫态或虫龄发生峰日相距防治适期的天数进行预警。
  - 物候法：利用其他生物现象作为害虫发生期的预警指标。
  - 有效积温法：每种昆虫完成某一发育阶段均需积累一定量的有效温度，即有效积温(K)。当环境温度(T)高于昆虫的发育起点温度(C)时，昆虫开始发育，通常有效温度($T-C$)与完成某发育阶段所需的时间(D)成反比，其关系式：$K=D/(T-C)$。
- 发生量预警：预警害虫的发生数量或田间虫口密度，估计害虫未来的虫口数量是否有大发生的趋势和是否会达到防治指标。但需要坚持多年积累有关资料，预警结果才具有可靠性。
- 灾害程度预警：在发生期、发生量等预警的基础上，根据作物栽培和害虫猖獗相结合的观点，进一步研究预警某种作物对虫害最敏感的时期，是否完全与害虫破坏力或侵入力最强而且虫量愈来愈多的时期相遇，从而推断虫灾程度的轻重或所造成损失的大小；配合发生量预警进一步划分防治对象田块，确定防治次数，并选择合适的防治方法，以争取治虫的主动权。

(3) 根据预警时效，可将预警分为短期预警、中期预警、长期预警和超长期预警

- 短期预警：对病虫害几天以后的发生动态做出预警，一般为 7～10 天，也称为“紧急预警”。
- 中期预警：对病虫害一个月以上的发生动态做出预警，一般为 10～30 天，也称为“警报”。
- 长期预警：对病虫害几个月的发生动态做出预警，一般为 30 天以上。
- 超长期预警：在研究病虫害发生发展规律的基础上，探索下年度的发生趋势，并以此为依据做出超长期预警。

3.2 农业病虫害预警指标体系

3.2.1 预警指标筛选原则

预警指标的筛选应遵循以下原则。

(1) 科学性。病虫害预警系统指标的选择和设计应建立在科学的基础上，应能充分反映病虫害发生发展的内在规律，指标的定义、计算方法等都要有科学依据，具有真实性和客观性。

(2) 可操作性。病虫害发生发展的机理复杂，影响因素繁多，在构建病虫害预警系统的指标体系时要考虑数据获得方便、计算简便等特点，使所构建的系统指标体系具有较强的可操作性。

(3) 准确灵敏。这是指病虫害预警指标与病虫害发生发展的态势具有较高的关联度，指



标的变化预示着病害或虫害的发生，具有较高的灵敏度，即预警指标应能够准确、灵敏、迅速、及时地提供预警信号。

(4) 简捷可靠。简捷是指预警指标的选择既要全面又要避免繁杂，能够揭示警情的变化规律；可靠是指预警指标数据的权威性和统计口径的一致性，统计数据的样本数量足够大，也就是需要有较大的时间序列，以满足预警和不断调整的需要。

(5) 互相匹配。这是指预警指标要与具体的预警互相匹配，也就是警素不同，预警指标也各不相同；互相匹配还指不同的病虫害要与不同的预警方法互相匹配。

(6) 稳定性。指预警指标在病虫害的变化过程中，对相对稳定的先行、同步和滞后的时间发生变化，而不是表现出一时先行、一时滞后、一时同步，否则就很难找出其中的规律。

3.2.2　预警指标体系的内容

针对农业病虫害中的种群分布、空间分布及其时间分布特性，可将农业病虫害的预警指标体系主要分为警源指标、警兆指标和警情指标，它们密切相关。警情产生于警源，警源只有经过一系列的量变与质变过程，才能导致警情的暴发；警情在暴发之前总有一定的先兆即警兆的出现；根据警兆指标的变化状况，联系警兆的报警区间，参照警素的警限确定警度。农业病虫害预警系统要以警情指标为对象，以警源指标为依据，以警兆指标为主体。其中警源指标以传染源和环境因素为主，警兆指标以疾病症状为核心，警情指标以宏观表现(发病个体、死亡率、发病率、空间布局)为核心。以动物病害预警为例，预警指标体系如图3-3所示。

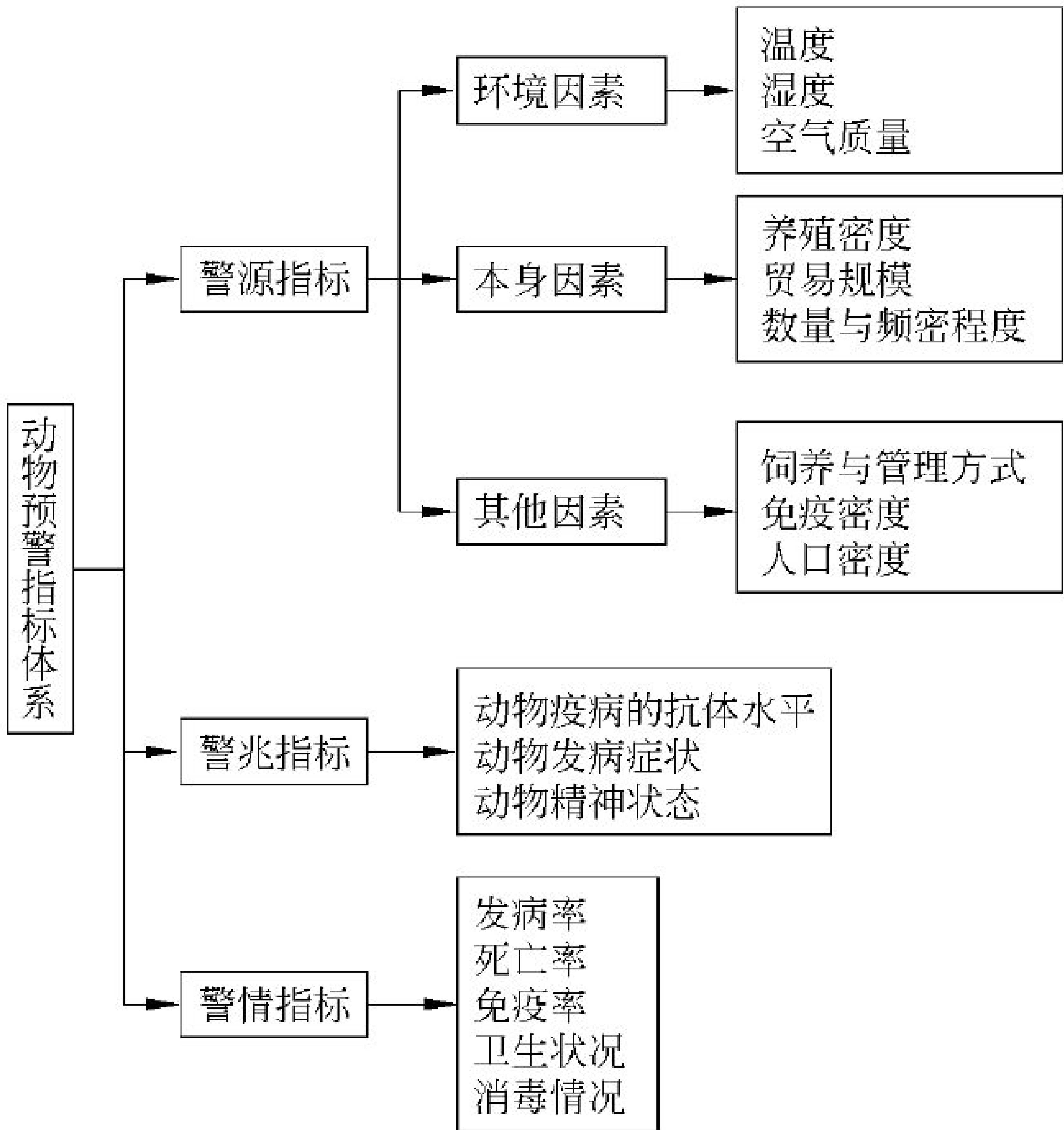


图3-3　动物预警指标体系

(1) 警源指标

警源是指警情产生的根源，用来描述和刻画警源的统计指标称作警源指标。从警源的生成机制看，警源指标可以分为两类：一类是农产品本身因素的警源即自生警源指标；另一类是由外部输入的警源，比如温度指标、湿度指标和空气质量指标等。寻找警源既是分析警兆的基

础，也是排除警患的前提。不同警素的警源指标各不相同，即使同一警素，在不同的时空范围内，警源指标也不相同。因此我们必须针对具体警素，寻根究底，顺藤摸瓜，直至找到问题的症结。在农作物预警、畜禽预警、水产预警中，根据研究对象的不同，会有不同的警源指标。以农作物预警为例，警源指标为环境因素、管理因素和其他因素。气候因素包括光照指数、气温、湿度、土壤污染程度、降雨量等；管理因素包括施肥次数、种植密度、浇水次数、人员管理水平等；其他因素包括农作物的种植时间、当前季节、种植坡度、农作物的品种等。

(2) 警兆指标

警兆是指警素发生异常变化导致警情暴发之前出现的先兆。用来描述和刻画警兆的统计指标称作警兆指标。一般不同的警素对应着不同的警兆，相同的警素在特定时空条件下也可能表现出不同的警兆，因此必须建立起合理的警兆指标。警兆指标又称先导指标或先行指标，它是预警指标的主体，是唯一能够直接提供预警信号的一类预警指标。所谓警兆识别就是对预警指标进行时差分类、筛选，确定其中的先行指标。

(3) 警情指标

警情指标是预警系统研究对象的描述指标。警情是事物发展过程中出现的异常情况，也就是业已存在或将来可能出现的各种各样的问题。用来描述和刻画警情的统计指标称作警情指标。警情的严重程度可以采用警级来描述，一般分为五级，即无警、轻警、中警、重警和巨警，分别用蓝、绿、黄、橙、红信号灯来表示。

3.3 农业病虫害预警系统

3.3.1 农业病虫害预警系统功能

针对农业病虫害的预警需求，农业病虫害预警系统主要应包括数据采集管理、指标维护评价、病虫害预测预报、病虫害预警分析、病虫害防治决策五方面的功能，[4,5] 如图 3-4 所示。

- 数据采集管理：主要指监测、采集、管理数据。如果缺乏统一的数据采集标准，数据的描述和表示各不相同，就会造成数据难以共享。
- 指标维护评价：指标是预测预警工作的关键，对于同一项预测或者预警，若选取的指标不同，其预测预警结果也不同；随着人们对病虫害发生机理研究的深入，其相应的预警指标体系也会发生变化，只有这样，才能保证预测预警的准确性。
- 病虫害预测预报：系统汇集各种人工神经网络、支持向量机、组合预测等模型。
- 病虫害预警分析：对警情进行趋势分析，报告警情范围。
- 病虫害防治决策：根据适时防治、科学防治的要求提出防治决策，指导农业生产。

3.3.2 农业病虫害预警系统结构

为了满足农业病虫害的预警功能，农业病虫害预警预报系统应该具有数据采集管理模块、指标评价知识模块、农业病虫害预警模块、预警预报模块及防治决策模块。系统模块构成如图 3-5 所示。



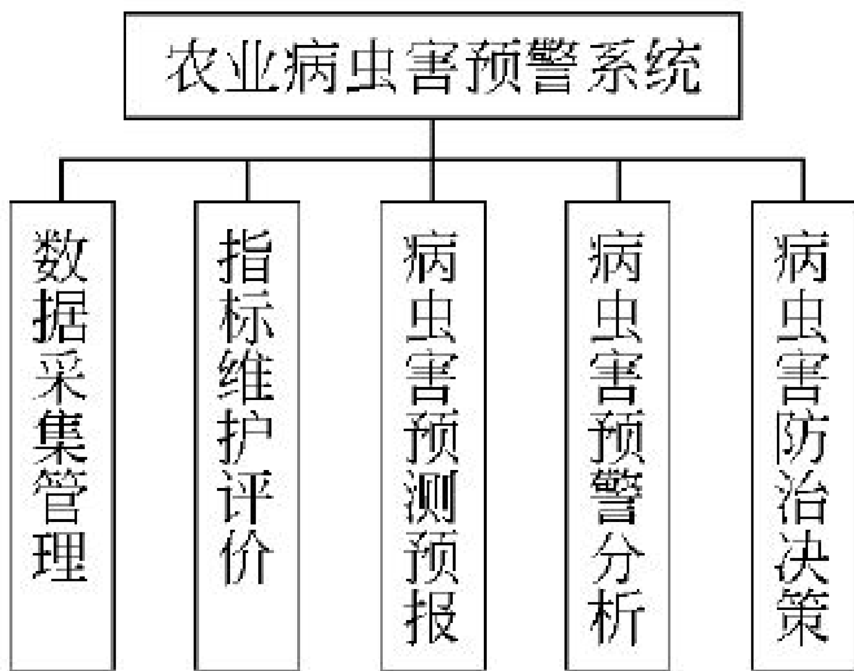


图 3-4　农业病虫害预警系统功能

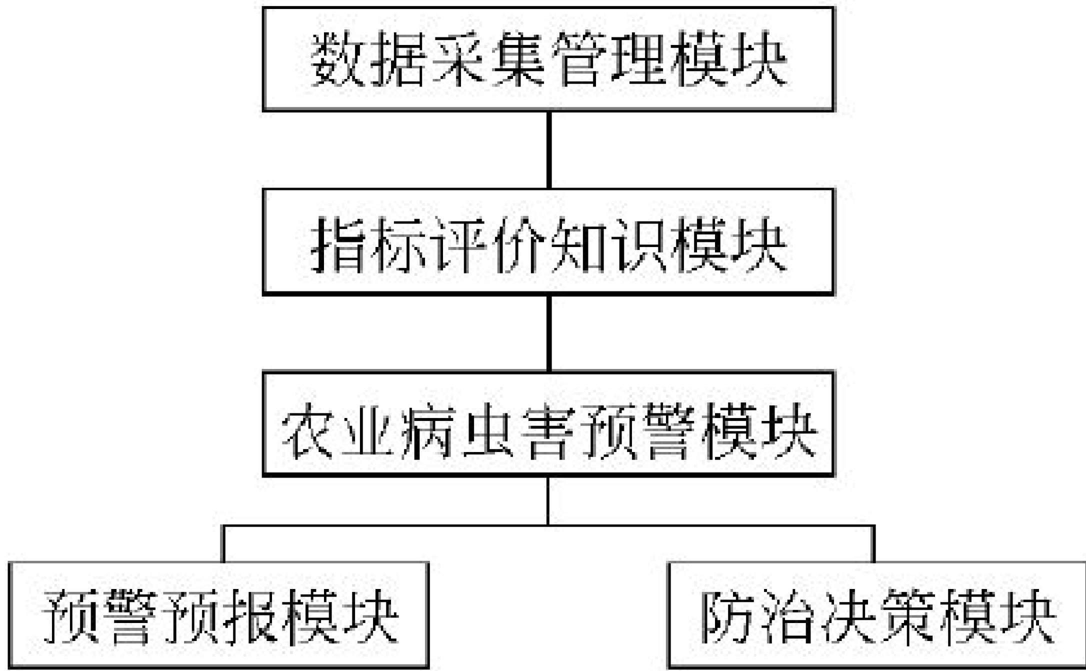


图 3-5　农业病虫害预警系统模块构成

3.3.3　农业病虫害预警系统的构建过程

农业病虫害预警系统的构建过程可参考图 3-6。

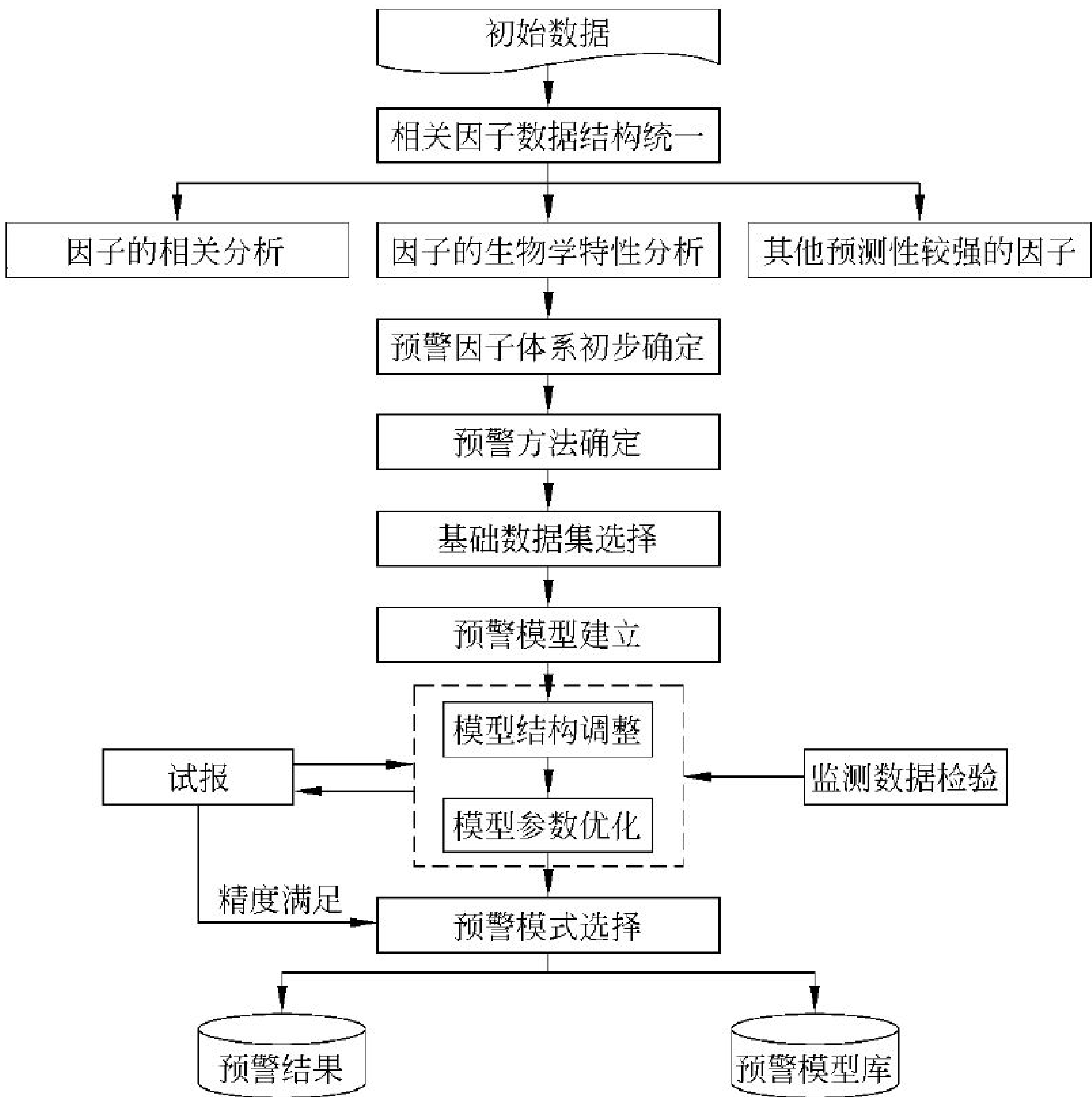


图 3-6　农业病虫害预警系统构建过程

3.4　小结

本章在分析研究农业病虫害预警内涵、过程、方法和特点的基础上，分析了农业病虫害预警指标体系的构建原则和构建过程，阐述了病虫害预警系统的功能和构建流程。本章的理论方法和技术路线为后续预警系统的构建奠定了基础。

参考文献

1. 李丽.基于径向基网络和支持向量机的梨病虫害预警预报系统研究[D].中国农业大学,2007
2. 肖悦岩.病虫害监测与预测[J].植保技术与推广,2002,22(4):40~41
3. 张孝羲,李运臂.昆虫生态及预测预报.北京:农业出版社,1985,300~333
4. 王骞.网络苹果无公害病虫防治专家系统研制[D].西北农林科技大学,2004
5. 乔晓亮.洛川无公害苹果主要害虫监测系统的初建[D].西北农林科技大学,2005

第 4 章

基于群决策的病虫害会诊理论[①]

基于群决策理论的农业病虫害远程会诊，可以辅助专家实现信息、经验和知识的共享，解决病虫害诊断系统存在脆弱性、柔性差和对于复杂病症难以诊断等问题，实现诊断决策的科学化、准确化。本章从群决策系统的角度对农业病虫害远程会诊问题的内容、过程、要素、组织结构以及方案选择等进行了分析，提出了远程会诊群决策问题的概念框架，并结合鱼病远程会诊进行了说明，为决策模型和方法的选择及构建提供依据。

4.1 病虫害群决策的内容和过程

农业病虫害的远程会诊是一个群决策过程，是网络化远程诊断决策过程中最重要的工作环节（见图 4-1）。农业病虫害远程会诊决策是远程诊断系统服务功能最高层次的应用，它以基于网络的农业病虫害诊断专家系统为主体，采用多种手段来收集病体本身的各种信息，以功能强大的系统软件与数学模型来完成周围环境理化指标的采集、监测和分析等任务，然后结合各种决策支持技术和方法，以达到对农业病虫害远程会诊过程中的各种群决策方案的产生和确定提供内容支持和过程支持。

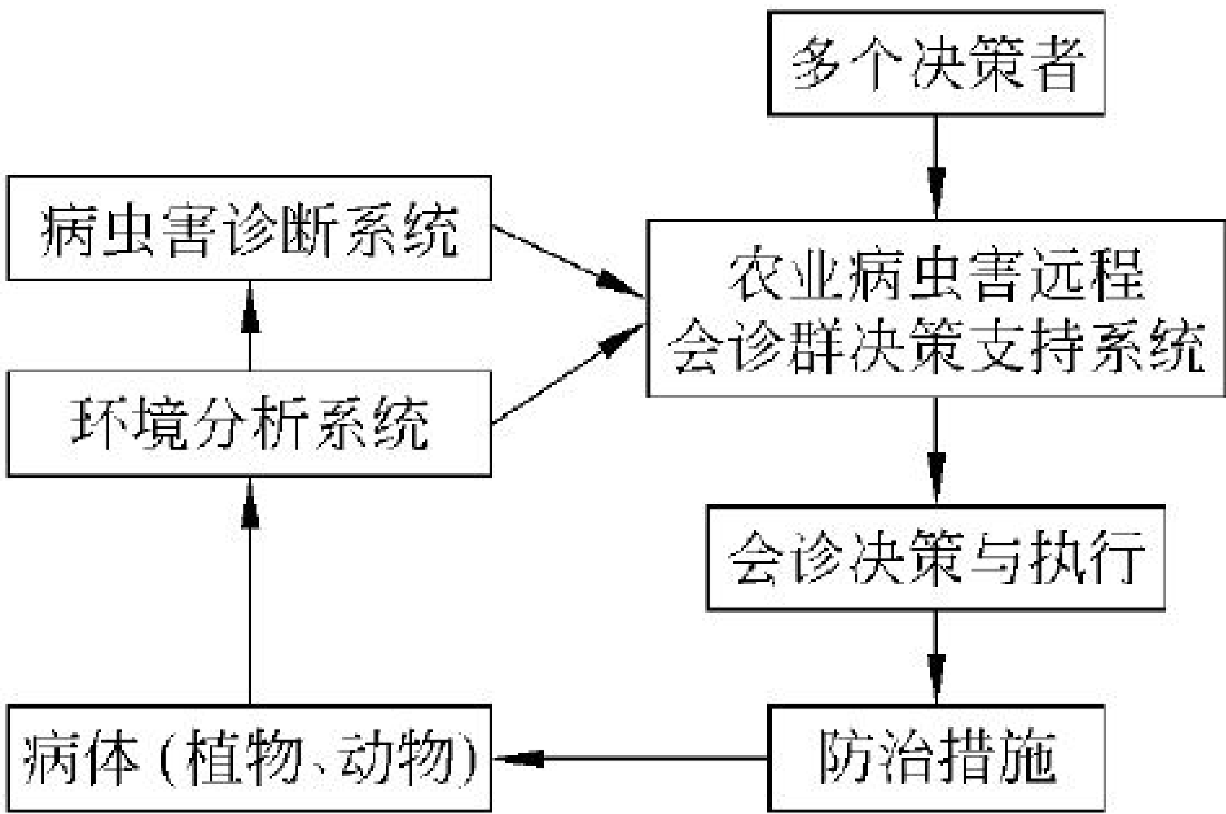


图 4-1　远程会诊群决策支持系统的定位

4.1.1 群决策内容

群决策内容支持旨在一定程度上能为用户提供探查特定领域问题实质的支持，体现为与决策问题有关的数据、信息、知识的处理，着重于问题的内容，试图在给定的社会或群体约束和目标下，找到最优解或满意解。

以鱼病诊断为例，鱼病的产生是外界环境的各种致病因素和鱼体自身抗病因素相互斗争，导致平衡破坏的结果。鱼类的致病因素复杂繁多，既有内因又有外因；既有直接致病因素，又有间接诱发因素；既有自然环境因素，又有人为饲养管理因素。在实际诊断中不仅要分析鱼体症状，而且还要利用水体的理化指标和微生物含量等信息对鱼病进行分析，这样才能使鱼病诊断信息达到全面、一致性要求，避免了在分析鱼病时仅仅依靠症状作出决策的狭隘性。鱼病远程会诊群决策内容支持就是依靠各种设备和手段，如数码相机、扫描仪等，综合分析现场调查、

① 本章选自研究人员杨萍的博士论文“鱼病远程会诊群决策支持系统研究”。

鱼体检查、检验诊断中所获取的诊断资料，然后将这些诊断信息分类（见图 4-2）存入各种类型的数据库，如病历库、图像库等。

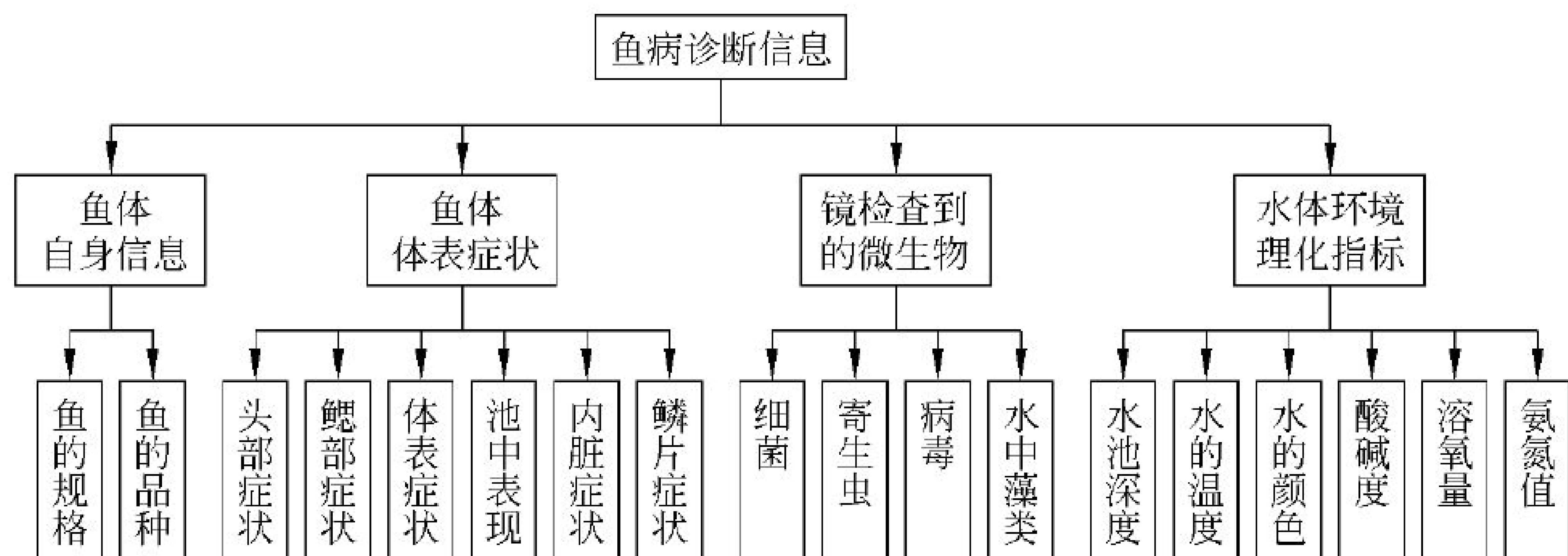


图 4-2　鱼病诊断信息分类

图 4-2 中的相关概念解释如下。

鱼体自身信息：鱼的规格、品种等。

鱼体体表症状：包括体表、头部、鳃部、鳞片和内脏五个部位以及鱼体的池中表现等诊断信息。

镜检查到的微生物：指对水体和鱼体表进行镜检查到的微生物，有病毒、细菌、藻类与寄生虫等。

水体环境理化指标：水体的溶氧量、酸碱度、氨氮值、水深和水温、水色。

4.1.2　群决策过程

群决策过程支持旨在一定程度上能支持或影响群体会议的进程，体现为成员参与和信息交换模式的改变，群体成员行动、态度、信念以及明显或隐含的规程的调整。[1]过程支持分为通信支持、过程结构化支持和智能过程支持三个等级，其中智能过程支持指开发基于知识的部件，为群体通信和过程结构化提供支持。

群决策往往是一个带有极大启发性的过程，由于有若干群体成员的参与，他们各自的资历、智慧和经验都各不相同，各人获得的信息也不相同，因此，在选择备选决策方案时，会产生不同的理解，表现出不同的偏好，作出不同的抉择。相对于个体决策而言，群决策更适合于半结构化和非结构化问题的决策，这些问题一般具有多种可供选择的方案，并且关于客观环境的信息不确定，以至于方案的预期效果及可能性很难确定，明确识别或选择最优方案十分困难。

这时，重要的工作就是要在成员中交换信息，沟通思想，加强理解。因此与通常的决策过程不同，通过沟通解决决策判断的分歧。沟通是一个交换信息、增进理解、集思广益的过程，会议是常用的沟通方式，但更重要的是非正式形式的沟通，一般要经过下列步骤：

① 提出问题；

② 讨论并探索不同的方案；

③ 各人对方案进行评价；

④ 集结群体的意见并检验是否满足群体的一致；

⑤ 若满足群体择案规则的一致性，则决策结束，否则返回第②步。

由于每一个成员都参与决策，并对决策结果各有不同的期望和偏好，而最终需要选定一个决策方案，这必然带来分歧和冲突。因此，必须事先确定选择备选方案的规则。典型的择案规



则大致有以下 4 种，即完全一致、商议一致、多数通过和等级决定程序。一般使用“多数通过”原则。

农业病虫害远程会诊群决策支持的过程支持就是通过视频会议系统和决策支持系统实现对决策任务、任务背景的全面交互和共享分析，相互协商产生各种治疗方案，以便及时达成一致满意的确诊治疗方案，如图 4-3 所示。

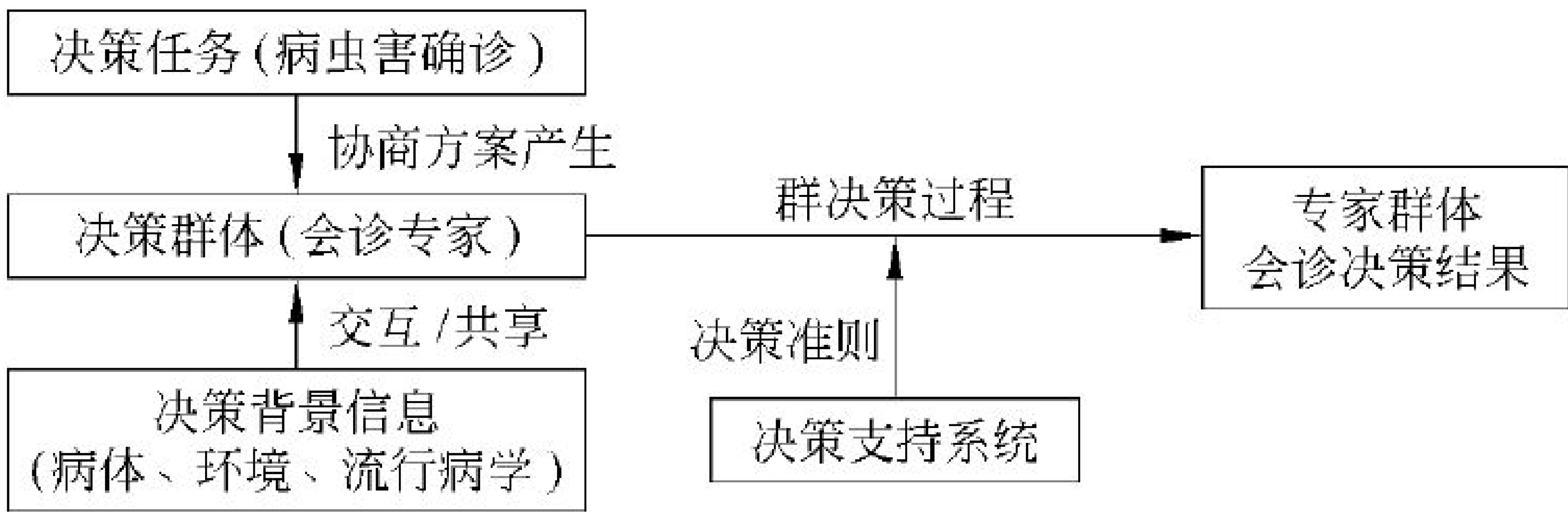


图 4-3　农业病虫害远程会诊群决策系统的过程

4.1.3　群决策特点

群决策有以下特点。

（1）会诊决策的时间、任务和成员的不确定性。会诊时间有一定随机性，仅当遇到复杂病症时才进行会诊。对不同病害的会诊，参与人员及其任务具有不确定性，由于复杂病害的不可重复性，每次会诊都是针对不同病情以及各个专家的专业特长来组成会诊专家群体。

（2）会诊决策成员的分布性。由于针对某种复杂病症，当地病虫害医生凭自身的知识经验无法确诊，借助远程会诊系统，他们可以预约各地同行专家共享信息，参加讨论，参与表决等。根据会诊的级别和规模，会诊决策成员可能来自不同省、市甚至不同国家，因而在地域上具有分布性。

（3）会诊决策过程的不确定性。会诊群决策过程是一个动态过程，其间会存在许多不确定性因素，如会诊成员相互之间可以相互商讨、提问等。在同一会诊中，决策成员在同一会诊中心的不同会诊议题或同一会诊议题的不同会诊中心中扮演的角色也会不同，存在不确定性。

（4）会诊决策过程的并发、并行和异步性。并发是指多个事件在某种条件或者状态下有可能同时发生。在农业病虫害会诊群决策过程中，可能存在多个决策成员同时抢占某种资源（如话筒、网络等）的情况。决策成员独立进行方案选择，具有并行特征。同时，由于农业病虫害远程会诊是一个分阶段的过程，因此还具有异步性。

综上所述，会诊决策成员的分布性决定了农业病虫害远程会诊决策过程具有群决策的特征，决策成员主要通过网络进行决策；根据会诊决策过程的并发性、并行性、异步性，利用群决策技术支持辅助会诊决策过程可实现会诊过程的各个方面的协调化和科学化。

4.2　群决策系统分析

4.2.1　群决策系统要素

诊断决策的理论和方法研究大多是从个人角度出发的，这是指诊断者对于诊断对象，从其症状表现出发，根据诊断知识作出诊断结论的过程这样一个思维推理过程。但是随着决策问题复杂性逐渐增大，决策过程不仅受到环境信息、个人偏好、方案评价方法等因素的影响，同时

问题的多目标性、不确定性、时间动态性和竞争性，也使得个人能力在决策过程中已远远达不到要求，需要集中多人的能力来完成诊断任务，于是推动了个人决策必然向群决策大量转移。这种有多个决策人参与决策分析、共同制定决策的整个过程就称为群决策，参加决策的人称为决策群体。[2]例如，鱼病会诊就是一个群决策。由于鱼体系统和生活水体环境的复杂性，对于某些复杂病例，一个鱼病医生可能没有足够的知识来作出诊断，此时需要专家之间的交互与协作，以共同完成对复杂病例的诊断和治疗。

群决策是一个系统，它的构成要素有：决策者(协调者)、决策成员、决策支持系统。决策成员通过相互交流和共享信息，按照某种决策规则确定或评价决策方案。其中，决策者可以是提出会诊申请的本地农业病虫害医生，也可以是农户。决策成员则包括参与会诊的所有农业病虫害医生和专家，决策支持系统是基于 Internet 的农业病虫害远程会诊群决策支持系统。群决策过程是包括数量集结和行为集结的农业病虫害远程会诊群决策过程。

4.2.2 群决策系统结构

群决策方式能给选择过程带来一些好处：更广泛的知识和经验，更多样化的视角，潜在的协作行为。同时也会带来一些不利之处，如果不能充分重视这些不利因素，则可能导致有问题的决策，甚至产生严重的后果。过多的决策参与者可能导致较差的决策，或者干脆没有结果。许多高级管理者厌倦了小组会议，竭力避免这种会议。大多数情况下，问题并不在于有多少人制定决策或参与制定决策，而在于参与者的组织结构是否适合具体决策问题。

1991 年 Holsapple 把多人参与的决策定义为“一种集合体的行为，这种集合体由两个或两个以上的个体组成，并且其性质由集合体的公共属性和组成成员的个体属性共同决定”[3]，根据这个定义，1997 年 Marakas 和 Wu 对“多人参与的决策”进行了分类[4]，分类如图 4-4 所示。

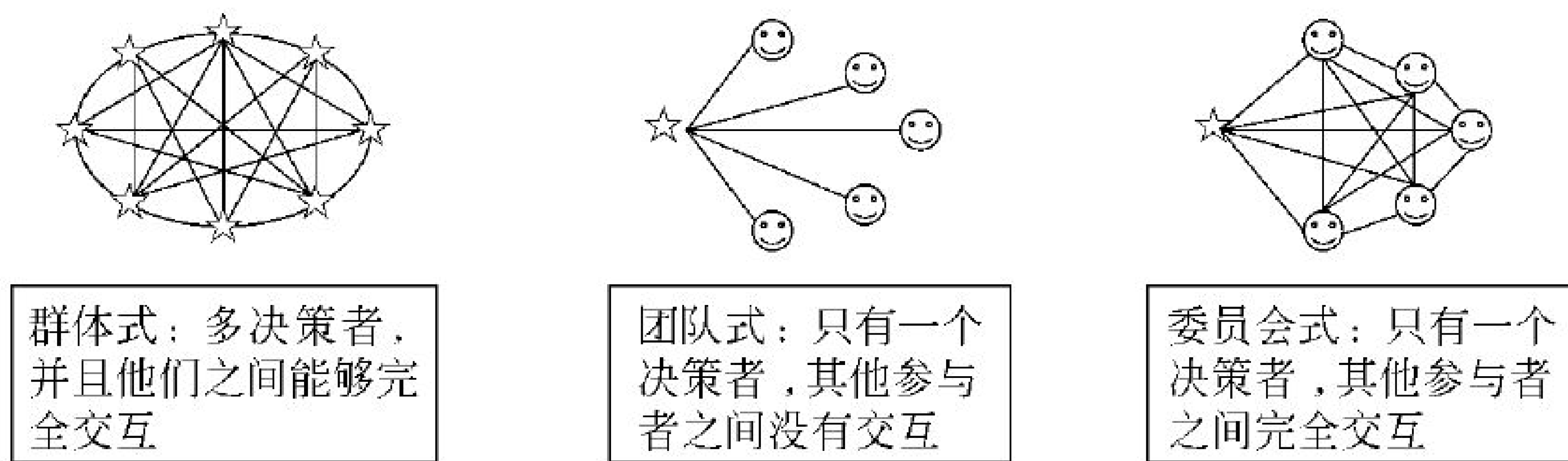


图 4-4　多人参与决策群体的基本结构

使用这种分类方法，易于确定多人参与的决策属于何种结构，并且搞清楚它是否适合解决某个给定的问题。每一种类型在决策者以及参与者之间的相互作用上都有特定的结构。

委员会决策的基本特点是各成员之间地位平等，原则上各成员的权利基本相同。根据委员会基本职能的不同，又可分为两种：一种是作为权力机构的委员会；另一种是作为参谋班子或咨询机构的委员会。前一种委员会本身就是决策实体，其中不再有最终决策权领导人，这种委员会的基本功能是要集结各成员的偏好以形成群(即社会)的偏好，得到最终决策。研究这种问题多采用社会选择理论，包括投票表决体制、社会选择函数和社会福利函数等方法；后一种委员会称为专家咨询及群体(公众)参与，其中各成员的权利虽然平等，但其作用是提供各成员的经验、知识和信息，帮助决策人作价值判断，最终决策往往由组建委员会的高级行政人员作出。

以上两种委员会决策问题的另一个重要区别在于前者往往有现成的备选方案集；而后者



则通常要在研究问题的过程中形成方案，再给出对各方案的评价意见。对于模糊决策树(FDT)群决策问题来说，作为会诊服务的客户方，也就是申请会诊的农户或者本地医生，遇到难以解决的病虫害问题时，通过FDT_GDSS向异地专家申请会诊服务，但最终诊断决策方案的确定和执行还应该是他们自己。因此针对远程会诊的特点，此处的群决策结构采用专家咨询式委员会结构。会诊专家只是决策参与者，专家群体一起对病虫害进行研究，给出初诊方案集，再对方案进行评价，提出确诊建议，而最终防治方案应由直接负责的本地医生结合具体情况和自身经验来制定。

4.2.3 群决策系统分类

依据不同的标准或参照不同的目标，有以下不同的群决策系统分类方法。

(1) 从群体成员之间的利益关系出发，可以将群决策划分为两大类：一类是各成员追求自身利益最大化，成员间存在利益冲突，这类问题常用对策论来解决。另一类是各成员均以追求群体的整体利益最大化为目标，在利益上一致。农业病虫害远程会诊就是第二类群决策，其特点是群体有共同的利益和目标，各成员除认识和经验上的差别外，不存在根本的利益冲突，也称为合作型群决策。一般的专家评审会、决策委员会、董事会、股东大会等，也都属于第二类群决策的组织形式。

(2) 按群体中各成员间有无“交互”，可将群决策划分为有通信的群决策和无通信的群决策。通信的目的是为了改变决策者的偏好。在群决策初期，每个决策者的优先选择未必相容，通过相互讨论，引导决策者改变态度，可以达成群体意见的一致。无通信的群决策的优点是决策过程快捷有效，但各成员之间的意见分歧未能得到解决。农业病虫害远程会诊属于有通信的群决策，我们研究的基于Internet的FDT_GDSS可以为同地或异地专家提供一个交互式讨论环境，实现专家之间知识、经验和信息的共享，提高会诊效果和质量。

(3) 按集结内容的不同，可以将群决策划分为三类：第一类是个体偏好的集结，通过集结个体对方案集的判断结果(偏好)得到群的偏好。第二类是个体不确定信息的集结，包括概率集结和模糊集结。这两类方法为数量集结，属于内容导向型方法。第三类是个体行为的集结，通过对各成员交互行为和反馈行为的描述，从个体行为集合得到群体的结果，属于过程导向型方法。

当然，某个群决策问题可能是不同分类的组合，而且经研究发现，采用集结内容不同的群决策分类方法有利于群决策方法的研究。在农业病虫害远程会诊群决策中，既存在成员之间的交互，又是通过对个体偏好的集结来得到群的决策结果的。

4.2.4 多属性群决策过程

多属性群决策希望解决的问题是集结群体成员的偏好以形成群的偏好，然后根据群的偏好对决策方案进行排序或从中选择群所最偏爱的方案。[5]

多属性决策的整个过程可以分为构造属性集(评价指标体系)、确定属性值、确定属性权重、综合排序几个阶段。

(1) 构造属性集

属性集的构造是多属性决策过程的起点，对于给定的决策方案集，用什么属性来反映它们的本质及差异，直接影响到多属性决策的后面几个阶段及最终决策结果。属性集的确定一般都涉及很强的专业知识和技能，单靠一个决策者很难获得满意的属性集，因此需要多位决策者



（专家）参与，最终确定出大家认可的满意属性集。群体构造属性集可以采用德尔菲法和名义群体技术（nominal group technique）等方法。

（2）确定属性值

为了构造决策矩阵，必须确定各决策方案在不同属性下的属性值。对于定量属性，单个决策者根据有关资料即可给出各决策方案的属性值，但对于定性属性，由于其本身的不确定性和决策者判断的模糊性，单靠一个决策者很难给出正确和客观的属性值，比较科学的做法是邀请多位专家分别给出各方案的属性值，然后利用适当的集成方法（如平均值法）获得相对准确的属性值。

（3）确定属性权重

属性赋权是多属性决策中的一个难点。对于主观赋权法，由于决策者知识、经验和偏好的差异，不同决策者作出的判断存在差别，所以不能只凭一个决策者的判断来确定权重向量，而是应该综合多个决策者的判断，得到一个综合权重向量，这就是一个群决策问题。例如，利用层次分析法（AHP）或模糊层次分析法（FAHP）确定权重向量时，不同决策者给出的判断矩阵一般是互不相同的，如何综合多个判断矩阵的信息以获得综合权重向量，是基于 AHP 或 FAHP 的群决策方法所要解决的问题。

（4）综合排序

对于给定的多个决策方案，分别由多个决策者进行决策，给出决策结果，即对方案的偏好排序结果，然后采用某种集结技术将这些排序结果集结为群体排序，即将个体偏好集结为群体偏好，这是涉及整个多属性决策过程的群决策。这种多属性群决策可分为两种：一种是从确定属性集开始就由各决策者单独进行决策，因此各决策者使用的属性集一般是不相同的，从而决策矩阵、属性权重也互不相同，排序方法可能相同，也可能不相同；另一种是各决策者面对同一属性集和同一决策矩阵，各自单独进行决策，然后再综合他们的决策结果。

总之，基于综合排序结果的群决策方法所要解决的问题是寻找群体一致性算法。以鱼病会诊为例，考虑到知识获取方面的困难以及前期的研究基础，初诊方案集由各专家依据患病鱼体信息，通过鱼病诊断专家系统并结合自己的知识和经验来获得，群决策专家只是针对方案集给出自己的偏好信息，选用合适的偏好集结方法来获取确诊方案。

通过实地调研得知，在鱼病诊断实践中，由于鱼体生活在水中，病情影响因子很多，一般在患病初期是难以发现的。等到发现时，情况基本都是比较严重，况且并发症居多。而现有的鱼病诊断专家系统的求解策略是输入鱼病信息后，要么得出一种病，要么无解，对于并发症是无能为力的。在初诊环节，各专家凭借自己的经验以及专家系统的建议，给出初诊方案集，方案集中的方案可以是并发病，再度处理时并发症也作为方案集中的一个方案来进行分析处理。因此，一方面提升了专家的主观能动性，充分利用了专家的经验和创造性，另一方面降低了知识获取的难度，使得会诊诊断更加接近实际，从而实现了定性分析（专家系统以及专家本身）和定量分析（决策模型）相结合的会诊决策求解策略，提高了系统的实用性和灵活性。

一般地，群体中成员的个人意见不可能毫无差异，要使群体集结的意见尽可能一致，除保存相同意见外，还必须通过中间人进行成员间的协商，通过协调、修正不同意见，尽可能减少成员间的意见差异，这个过程称为一致性达成过程。发展这个过程的协调技术，促成计算机充当中间人进行人工智能协调，是非常有意义的。



4.2.5　群决策规则

作决策时,决策者总是力图选择“最好的”可行方案,这就需要对方案根据其所有属性值排列优劣次序(或分档定级),而对方案排序或分档定级的依据称作决策规则。

决策规则可分为两类:一类是最优化(optimizing)规则,它能集中方案,把所有备选方案排成完全序;而根据决策规则所蕴涵的某种准则,在完全序中总存在一个最好的方案。另一类是满意(satisfying)规则,它为了使分析简化、节省时间、降低费用而牺牲最优性,把方案集划分成容易处理的几个有序子集,比如可接受与不可接受两个子集,或者好、可接受与不可接受三个子集;根据这种规则,不同子集里的两个方案的优劣是显而易见的,同一子集中的方案无法或难以分辨优劣。

(1) 一致性规则

由于决策群体中各个决策者之间存在利益或意见冲突,要想得到一个统一的决策结果,群体只能寻求妥协或一致。任何一个决策群体,如果没有寻求一致作为内在固有的、相对稳定的基本特性,则不可能找到某种被全体成员所接受的并共同遵从的客观规则。实际中,群体关心的往往不是作出什么决定,而是如何就最后的决定达成一致。因此,群决策的准则之一是极大化一致性。

必须看到,追求群体一致性可能会抑制群体的一些创造性思想,而群决策的最大优势是它可能比个体决策存在更多的创造性思想,如何权衡二者是研究群决策一致性要考虑的问题。

(2) 效用极大化规则

在个体决策中,一个理性的决策者总是使决策的期望效用最大化,群决策也是期望群体效用最大化。

设 m 个决策者构成决策群体,对方案集 $X=\{x_1,x_2,\cdots,x_n\}$ 进行决策,$u_i(x_j)$,$i=1,2,\cdots,m$,$j=1,2,\cdots,n$ 是第 i 个决策者对第 j 个方案的效用,群决策效用极大化准则为

$$\max\ \{u_i(x_j)\mid i=1,2,\cdots,m,j=1,2,\cdots,n,x_j\in X\}$$

对于农业病虫害远程会诊群决策问题来说,由于医疗诊断的复杂性和知识表达的局限性,难以进行价值判断,期望效用很难衡量。这类决策问题是目标取向的,其决策结果的正误与优劣可以客观地衡量,因而每个个体的方案选择及最终的集结结果一般有客观上的正误之分,这与社会选择的价值取向是不同的。因此,农业病虫害远程会诊群决策准则就是群体不断磋商,最后形成一个满意的相对一致的确诊方案。

4.3　群决策问题的求解

4.3.1　病虫害会诊群决策的目标导向

理性指的是一种行为方式,即在给定条件和约束的限度内适于达到给定目标的行为方式。我们把那类考虑到活动者信息处理能力限度的理论,称为有限理性论[6]。

传统经济理论假定了一种“经济人”,这种人在行动过程中既具有“经济”特征,同时也具有“理性”。传统经济理论认为,这种人具备关于其所处环境各个方面的知识,而且这些知识即使不是绝对完备的,至少也相当丰富和透彻。此外,这种人还被设想为具备一个很有条理的、稳

定的偏好体系，并拥有很强的计算技能；依靠这类技能就能计算出，在他的备选行动方案中，哪个方案可以达到其偏好尺度上的最高点。

理性抉择的古典模型要求决策者了解一切备选方案；要求完全了解每一方案所导致的后果和具备计算这些后果的本领；要求决策者现在和将来对这些后果的评价是确定的，无论那些后果种类多么繁杂、性质多么不同。

可是现实生活中人们经常面对的情况是：没有现成的备选方案，不得不搜索备选方案；由于计算能力有限以及外部世界的不确定性，人们对选取具体方案的后果仅仅有着非常不完全的了解；决策者不具备用以比较不同方案的普遍而一致的效用函数。这种情形下，决策者怎样作出决策？

西蒙指出，现代管理决策的两个基本假设是：①决策者必须考虑决策环境，希望达到一个满意的目标水平。②各种经济组织（企业、公司、银行等）是一个合作系统，组成它的各个团体也许会有不同的甚至是矛盾的目标，但是它们必须互相协调，共同对策。基于这两个基本假设，就可以自然地用多准则决策模型来描述现代经济管理的决策问题。

西蒙指出，决策者由于受到认识上的限制，所以不可能知道他们的决策所产生的全部后果；由于决策环境的日益复杂，决策因素的日益增多，决策者也不可能了解全部的决策方案究竟有多少；另外，由于时间、金钱、人力、物力和资料来源的限制，也不可能把所有的方案都拿来进行一一比较。因此，得到真正的最优解几乎是不可能的，也是不必要的。

但是，决策者可以预先规定一个满足原定目标的最低要求，然后寻找满足这些最低要求的方案。这就大大简化了决策过程。很明显，“满意解”模型要比“最优解”模型丰富得多，也更加简化和容易接受。由此可知，现代管理决策所追求的不是绝对意义的最优解，而是相对意义的满意解。关于决策制定过程的两种观点：“完全理性假设”与“有限理性假设”的区别可以从决策制定步骤来区分（见表 4-1）。

表 4-1　决策制定过程的两种观点：“完全理性假设”和“有限理性假设”[7]

决策制定步骤	完全理性假设	有限理性假设
① 提出问题	确定一个重要的、相关的组织问题	确定一个反映管理者利益和背景的、可见的问题
② 确定决策标准	确定所有的标准	确定有限的一套标准
③ 给标准分配权重	评价所有标准并据此对组织目标的重要性进行排序	建立一个简单的评价模型并对标准排序；决策者自身的利益强烈影响排序
④ 制定方案	创造性地制定广泛的各种方案	制定有限的一系列相似方案
⑤ 分析方案	依据决策标准和重要性评价所有方案；每一方案的结果均是已知的	从希望的解决方法出发，依据决策标准逐一评价方案
⑥ 选择方案	最大化决策：获得最高经济成果的方案（依据组织目标）	满意决策：寻找方案一直到发现一个满意的、充分的解决方法为止
⑦ 实施方案	由于决策是最大化单一的、明确的组织目标，所以所有组织成员将会接受此方案	政治和权力的考虑将会影响决策的接受和执行
⑧ 评价	依据最初的问题客观评价决策成果	只有消除评价者个人利益才能客观评价决策结果

资料来源：斯蒂芬·P.罗宾斯. 管理学[M]. 北京：中国人民大学出版社，1998，127



显然，农业病虫害会诊决策专家对于复杂病症的求解，是在众多解中寻求一个满意解。首先，由于病症的复杂性以及专家经验知识的不完备性，遇到复杂病症时，没有现成的备选方案，需要专家们一起会诊搜索备选方案；其次，由于复杂病症的影响因素的日益复杂，很难完全了解每个方案所导致的后果，需要不断的假设验证。最后，由于时间的限制，专家们不可能对列出的各个方案逐一比较，只能在假设验证过程中不断修正，以取得满意方案。

因此，我们研究的基本假设如下所述。

基本假设：有限理性的农业病虫害会诊决策者很难选择最佳方案，决策的目标是寻求一个满意方案。

根据上述分析，农业病虫害会诊决策者是在若干已知解中找出一个满意解。这就明确界定了我们的研究手段是采用多属性决策方法（即在已知的有限个可行解中寻找满意解），提高农业病虫害会诊决策效果，而不采用多目标决策方法（即在无限个解中寻找最优解）。

4.3.2　病虫害会诊群决策方案选择

一般来说，决策分析中存在着三个基本要素：一是可供选择的方案，在决策术语中称为策略，它是一个多元集合，在这一集合中蕴涵着所要选定的目标；二是一组给定的约束条件，策略的选择与目标的追求都必须以满足约束条件为前提；三是一个已知的效用函数，用以衡量每种策略的得失。

目前的文献中，存在多准则决策、多目标决策以及多属性决策三个概念混用的现象。从备选方案的角度对其进行区分，多准则决策包括两大部分，凡是决策空间是连续的、备选方案无限的称为多目标决策。凡是决策空间是离散的、有限的称为多属性决策。本质上多属性决策研究已知方案的评价选择问题，这类问题求解的核心是对各备选方案进行评价后排定各方案的优劣次序，再从中择优。而多目标决策研究未知方案的规划设计问题，求解这类问题的关键是向量优化，即数学规划问题。两者的比较见表 4-2。

表 4-2　两类多准则决策问题的对照表[8]

多属性决策问题		多目标决策问题
决策变量	离散型	连续型，$\boldsymbol{x}=(x_1,x_2,\cdots,x_N)$
方案集	$X=\{x_1,x_2,\cdots,x_m\}$	$X=\{\boldsymbol{x}\mid g_i(\boldsymbol{x})\leqslant 0,i=1,2,\cdots,m,x\in R^N\}$
属性集	$Y=\{y_1,y_2,\cdots,y_n\}$或$F=\{f_1,f_2,\cdots,f_n\}$	用目标函数 $f_j(x),j=1,2,\cdots,n$ 表示
决策形势	只包括分析评价，根据属性矩阵进行分析评价的主要目的是对方案排序	包括系统建模。由模型生成方案集；分析评价主要是求解多目标规划问题，要从非劣解集中获取偏好解

资料来源：岳超源.决策理论与方法[M].北京：科学出版社，2003，164

多属性决策依据问题的多个属性进行方案选优或排序，是决策分析的重要组成部分。一般来讲，多属性决策问题可以分为三种类型，即选择问题、排序问题和分类问题。[9] 给定方案集，选择问题是从中找出最满意的方案集；分类问题是将方案集中的每一个方案分派到预先定义的类型中去，一个方案被分派到一个适当的类型中取决于方案的内在价值。排序问题是在方案集上建立一个偏序关系或者全序关系。

农业病虫害会诊群决策问题，主要是针对复杂病症会诊决策而言。针对这类问题的会诊决策，从决策单元的角度看，因为一位专家无法确诊，故需要多位专家参与决策，显然大多数大

型会诊决策具有多人决策中的委员会式决策特点。从备选方案的角度看,鱼病会诊决策者会针对病症提出一个初诊有限方案集。因此,具有多准则决策中的多属性决策特点。

再者,由于群决策至少有两名决策者需要共同负责决策,决策者之间存在能力、利益、信息来源、价值观等方面的差异,决策时必然存在意见和偏好冲突。群决策与个体决策的本质区别也正在于群决策存在意见冲突。

当决策者之间仅存在信息来源差异时,可以通过成员之间的交互和沟通来消除。如果决策者之间存在能力差异或价值观差异,可以通过学习或培训的手段来提高决策者的素质和决策能力。而当决策者之间存在利益差异时,一般来说很难消除意见冲突,只能通过事先制定的"决策规则"来获得最终的群决策结果。

因此,本研究的第一个预定假设如下。

预定假设1:①农业病虫害会诊决策普遍存在多属性群决策特点,会诊决策者的意见和偏好存在冲突和不一致性。

4.3.3 病虫害会诊群决策环境特点

不确定性是决策分析中存在的普遍现象。根据经验可知,事件的不确定性有两种不同的表现形式。一种是事件是否发生的不确定性,即通常所说的随机性;另一种是事件本身状态的不确定性,称为模糊性。对前者而言,事件是否发生虽难预知,但事件本身的状态是清楚的。例如投掷一枚硬币,哪面朝上事先并不知道,但每次投掷结果绝不含糊,不是国徽朝上就是数字朝上,非此即彼,这种现象在数学上称为排中性。但对后者而言,问题不在于事件发生与否,而在于事件本身的状态不很分明,致使不同的人观察同一事件会有不同的感觉,因而得出不同结论。例如在鱼病会诊决策中,"水的颜色发黄"、鱼体"池中表现不活泼"、"体表发黑"等许许多多的概念,都具有外延不分明的特点。相对于明确的概念,我们称这样的概念为模糊概念。包含模糊概念的事件属于模糊事件。模糊事件的主要特性之一是不服从数学上的排中律,不是非此即彼,而是亦此亦彼。

既然模糊性是事物客观存在的一种属性,因此是可以描述的。经典数学是以精确为特征的,面对模糊问题无能为力。为了从根本上揭示模糊性自身的规律,著名的美国加利福尼亚大学控制论专家扎德(Zadeh)教授通过对被视为现代数学基础的集合论的考查,认识到传统数学与人脑思维的根本差别在于其处理模糊信息的能力。事实上,人脑的思维、判断和推理在大多数情况下都是基于非量化的或不精确的观测和规则作出的。1965年他在*Information and Control*杂志上发表了一篇开创性的论文"Fuzzy Sets",标志着模糊数学的诞生。[11]

模糊数学的诞生为模糊决策的发展提供了理论基础,人们利用这一理论对模糊意见集中决策、模糊二元对比决策、模糊综合评判决策、模糊多属性决策、模糊多属性群决策等问题进行研究。

从诊断学角度来看,任何一种诊断信息都是模糊的、不确定的。任何一类诊断对象,单用一方面信息来反映其状态行为都是不完整的。从决策环境角度看,农业病虫害会诊决策环境中,不仅涉及定量因素,而且还有大量定性因素。这些定性因素的处理涉及模糊因素。由于复杂病症的影响因素日益复杂以及由于时间的限制,专家们不可能对列出的各个方案逐一比较,往往给出一个可信度或者可能度,然后在假设验证过程中不断修正,以取得满意方案。因此,

① 预定假设是为了一个正在进行的研究而假定为真实的事情,使一个研究目标可以实现的假定就是一个预定假设。[10]



具有模糊多属性群决策的特点。

因此,本研究的第二个预定假设如下。

预定假设2:农业病虫害会诊决策环境(即自然状态和专家认识)具有模糊性,因而决策方案集合中的意见具有不确定性特点。

4.3.4 非结构化的病虫害会诊决策

西蒙教授按性质将决策问题分为结构化决策和非结构化决策。结构化决策(structured decision)指的是那些问题的本质和结构十分明确,解决这些问题的步骤已知,而且是经常重复发生的决策问题。非结构化决策(unstructured decision)指的是那些以前未曾出现过,或者其本质和结构复杂而难以确切了解的问题,而用以往解决问题的一些方法和步骤难以处理的决策问题。决策问题按结构化程度不同分为:结构化问题、半结构化问题和非结构化问题。而划分的原则是视问题能否用标准过程求解。对于通用标准过程不能求解的结构化问题,个人决策往往更有利,节省时间和人力。对于属于组织中带有战略性质的问题,非结构化程度高,采用群决策的质量一般比个人决策的质量高。

运用群体方式解决的决策问题有这样几个特点:

(1) 受潜意识影响较大,与决策者的经验有很大关系,如果决策失误则造成的浪费较大,并且在开始行动以后,发现决策不好要改变或挽救也比较困难,甚至不可能。

(2) 存在多种可供选择的方案。

(3) 关于客观环境的信息很少,以至方案的预期效果及可能性很难确定。

(4) 明确识别或选择最优方案十分困难,这主要是由于方案的多目标非劣解和个人的偏好各异性所造成的。

(5) 所选定方案的执行结果不能很好地得到反映,要经过长期实施后才能反馈结果。

(6) 为了确保决策的实施,需要组织中的成员一致认同问题。决策质量并不是选择个人方式还是群体方式的充分条件,能否为人们所接受才是至关重要的,因为一个很好的决策如果不能为人们所接受,则无法贯彻落实,那么决策也成为没有用处的空谈。

决策问题的类型不同,通常采取的决策方式也不同。一般对结构良好的问题采取程序化决策(programmed decision),而对结构不良的问题采取非程序决策(non-programmed decision)方式。表4-3表明不同决策类型对决策要素的影响。

表4-3 决策类型对决策要素的影响[14]

	决策主体	备选方案	不可控因素	决策结果
经营性结构化和独立的决策	可能是具有完全自主权的个人担任决策主体,因为决策是受限制的。容易理解且不可能影响其他决策	备选方案比较明确,但是数量可能较多	可能有一些不可控因素,因为决策的有界性,这些不可控因素大多有案可查	每个结果可能有多种属性,但是可能是"可预料到的"
战略性非结构化和非独立的决策	可能需要多人组成决策主体,因为决策对整个组织至关重要,对其会有不同的解释并且会产生广泛的影响	可供选择的方案并不很明确,数量不多,因为方案不是难以产生(非结构化)就是需要妥协(非独立)	可能存在很多不可控因素,其中一些甚至不为人知,大多数难以预料	每一结果可能有多种属性,它们都难以预料

对于结构不良问题的决策,由于决策问题的复杂性,一般由多人作出决策。根据前面的分

类,多人决策被分为群决策和主从决策两大类。群决策的优点是:信息更完整,方案更多,对方案的接受程度更大,以及更具合法性。多人决策常用的四种方法是:头脑风暴法、名义群体法、德尔菲法和电子会议。

进行农业病虫害会诊决策时,复杂病症的确诊问题就属于对结构不良问题的决策。以鱼病为例,鱼类的致病因素复杂繁多,既有内因又有外因;既有直接致病因素,又有间接诱发因素;既有自然环境因素,又有人为饲养管理因素。既然是复杂病症,就是专家医生以前没有遇到过的,无经验可循,诊断信息不完备。这是一个典型的半结构化问题。虽然复杂病症是一个非常复杂的问题,但是对于鱼病诊断已经有了很多定性的分析方法,例如温继文和郭永洪等的研究积累了很多可以借鉴和作为指导的知识和经验,采用群决策可以在给出初诊方案集的基础上,借鉴定量的多属性群决策理论和方法,群策群力,实现信息和经验知识的交流共享,从而提高诊治效率和效果,为鱼病防治的信息化和科学化奠定一定的研究基础。[12,13]

因此,本研究的第三个预定假设如下。

预定假设3:农业病虫害会诊决策大部分属于对结构不良问题(ill-structured problems)进行决策。

4.3.5 病虫害会诊群决策的诊断和治疗假设

根据上述预定假设1~3,可以得出以下三个诊断假设。

诊断假设1:现实中,农业病虫害诊断决策多属性问题普遍采用定性决策,定量方法运用较少,在一定程度上制约着农业病虫害诊断决策效果的提高①。

诊断假设2:现实中,不确定性信息的处理以及模糊决策方法因计算量大、比较抽象,应用较少,在一定程度上制约着农业病虫害会诊决策效果的提高。

诊断假设3:现实中,对结构不良问题的决策,由于信息化手段很低,尤其是农业病虫害会诊决策支持系统发展很慢,在很大程度上制约着农业病虫害会诊决策效果的提高。

若能在某种程度上对农业病虫害会诊决策中的三个诊断性假设所涉及的问题进行改进,那么可望提高会诊决策效果。也就是说,上述三个诊断假设构成了需要解决的主要问题。

为此,提出以下三个治疗假设。

治疗假设1:群决策理论和方法是处理多人参与决策的一种定量方法,可有效解决农业病虫害会诊决策的多人会诊决策问题,进而提高农业病虫害会诊决策效果②。

群决策理论可以说既是决策理论的前沿,也是决策理论最为薄弱的部分,虽然相关研究始于二百多年前,但它作为一个明确的概念被提出来仅仅三十来年,得到广泛的研究则是在20世纪80年代以后。尽管它受到越来越多国内外学者的关注,新的科研成果也不断涌现,然而,由于相关学科领域发展的限制,如多目标决策理论和方法尚未完善,行为科学的研究对决策者的行为还不能作出精确描述等,群决策理论总体上尚未形成一定的理论体系。就其研究目的而言,群决策是为了综合群组中各成员的意见以形成群的意见,并根据群的偏好排列出行动方案的优劣次序。因此,对群组中决策个体偏好信息集结的研究仍是群决策研

① 诊断假设是关于一个问题的原因的命题。例如,对发展中国家的"廉价食品"政策导致低下的农业生产率的推理就是诊断假设(埃思里奇,1998)。

② 治疗假设被用于解决问题。治疗假设常常与诊断假设相联系,它们为诊断假设已指明原因的问题提供可能使其解决的处方。治疗假设典型地在经验上是不可检验的,但它们可以通过推理按照证据的重要程度被接受或拒绝(埃思里奇,1998)。



究的焦点。

对于鱼病会诊决策而言，由于鱼体系统和生活水体环境的复杂性，在实际诊断中不仅要分析鱼体症状（定性指标），而且还要利用水体的理化指标和微生物含量（定量指标）等信息对鱼病进行分析。遇到复杂病例，一个鱼病医生往往没有足够的知识进行诊断，此时需要专家之间的交互与协作，共同完成对复杂病情的诊断。因此利用群决策方法来辅助鱼病会诊决策是十分有利的。在决策属性不完全或者指标体系不完善的情况下，直接用来构造决策群体对方案集的偏好，进而用多属性群决策集结方法来获得满意方案也是有效的。

治疗假设 2：基于证据理论的信息融合方法可以有效集结不确定性信息，并能解决冲突意见的合成问题，进而提高农业病虫害会诊群决策效果。

一般来说，随机性是一种外在因果的不确定性，而模糊性是一种内在结构的不确定性。从信息观点看，随机性只涉及信息的量，模糊性则关系到信息的含义。可以说，模糊性是一种比随机性更深刻的不确定性。在现实生活中，模糊性的存在比随机性的存在更为广泛。尤其在主观认识领域（如会诊决策），模糊性的作用比随机性的作用重要很多。

目前进行不确定性信息融合的数学工具主要有贝叶斯的概率理论、证据理论、模糊集论和神经网络等。从解决问题的角度来看，证据理论的一个重要优势就在于对多信息源不确定信息的表述和融合，以得到更为准确有效的信息。证据理论用集合来表示命题，将对命题的不确定性描述转化为对集合的不确定性描述，讨论的对象是目标集的幂集，在对证据信息的表示上，能够区分不确定和不知道的差异，对不确定性问题的描述更接近人的思维习惯，能较好地处理具有模糊和不确定信息的合成问题。近年来，成为人工智能和专家系统中描述和处理不确定性问题不可或缺的工具。

在实际农业病虫害会诊决策中，会诊专家在对方案集进行确诊决策时，常常涉及许多定性指标，需要处理大量的定性信息，其中不可避免地存在不确定信息和不完全信息。而又因为进行这种决策需要丰富的相关知识及实际经验，所以业内专家的意见是必需且重要的。但同时应该认识到，专家的决策意见也带有主观性，也会存在偏差，并且有时会出现含糊或不确信的认知状况。可见，在这类定性决策中，信息本身存在的不确定性以及会诊专家认知的不确定性，使得确诊结果的最终确认和可信性受到挑战。而基于证据理论的信息融合方法正是解决此类问题的一个有效手段。

治疗假设 3：通过决策支持系统解决结构不良问题（半结构化/非结构化问题），通过决策支持系统的人-机对话系统引导决策者对问题的表达和理解，通过电子会议和一致性分析技术改进群决策效果，可有效提高会诊决策效果。

群决策问题一般是非结构化的复杂决策问题。这些问题往往庞大而且复杂，仅凭单个决策者有限的知识和精力，难以作出令人满意的决策，需要集中群决策者集体的智慧才能创造性地解决问题。

决策支持系统（DSS）的主要特征就是辅助决策人员完成半结构化和非结构化的决策问题，这些问题很少或几乎得不到管理信息系统的支持。表 4-4 列出了 DSS 提供支持的一般类型。应用 DSS，不仅仅是为了获得一些正确、合理的决策结果，提高工作效率，还要改进决策工作，提高人对问题的了解和认识能力以及寻求解决方法的能力，进而提高工作效能。

表 4-4　DSS 提供支持的一般类型[15]

- 从多个角度研究决策问题的背景
- 生成多个高质量的可供考虑的备选方案
- 探讨并测试多个解决问题的战略
- 促进头脑风暴以及其他创新性的问题解决技术
- 对于特定问题背景,探讨多种分析方案
- 对不断弱化的偏见与不恰当的试探法进行指导和约简
- 增强决策者处理复杂问题的能力
- 改进决策者的反应时间
- 阻止未成熟的决策与方案选择
- 对大量不同的数据进行控制

资料来源：George M. Marakas 著,朱岩,肖勇波译.21 世纪的决策支持系统[M].北京：清华大学出版社,2002.41

群决策支持系统(GDSS)是信息、通信、计算机和决策技术的结合。GDSS 中用到的信息、通信技术包括电子信息、内联网、外联网、Internet、电子会议、存储和交换设备等;计算机技术包括多用户系统、第四代语言、数据库、数据分析、数据存储和修改、面向对象技术,分布式智能技术和机器学习等;决策支持技术包括议程设置、自动推理技术、决策模型方法(决策树、风险分析、预测方法等)、结构化群体方法(名义群体法和 德尔菲技术等)及指导群体讨论的规则等。鱼病会诊决策一般是同地或异地的多位专家对一种复杂病症共同决策,会诊过程中可以利用 GDSS 促进专家之间的交流,从而提高会诊决策的效果。

对于大型农业病虫害会诊决策,需要群决策。这些群体决策过程往往是：根据已有的材料(收集到诊断信息),根据群体成员(有经验的农业病虫害医生或者从事诊断研究的专家学者)各自的经验和智慧,通过一定的议程(如农业病虫害远程会诊),集中多数人的正确意见,作出决策。而 GDSS 可提供三个级别的决策支持：第一层次旨在减少或消除群决策中决策者之间的沟通障碍;第二层次是提供善于认识过程和系统动态的结构技术,决策分析建模和分析判断方法的选择技术;第三层次是将第一层与第二层的技术结合起来,引导决策者参与决策。

将上述分散的思想集成在一起,就可为提高农业病虫害会诊决策效果提出一个整体解决的思路：建立基于不同形式偏好信息和不确定性信息的专家意见集结模型的农业病虫害远程会诊群决策支持系统模型库,对群体意见进行综合;通过群决策支持系统的支持,建立基于一致性分析的农业病虫害远程会诊群决策过程模型,充分考虑决策群体的决策思想及经验,促进意见集结,改进结构不良问题的解决。

4.4　小结

本章在描述了群决策问题的内容和特点的基础上,对农业病虫害会诊群决策系统要素、结构、类型、过程和规则进行了分析。在此基础上,对群决策问题的目标导向、方案选择、环境特点以及非结构化进行分析,最后给出了群决策问题的求解策略,为本书后续的方法研究奠定基础。



参考文献

1. 段新生. 证据理论与决策、人工智能[M]. 北京:中国人民大学出版社,1993
2. 杨雷. 群体决策理论与应用一群体决策中的个体偏好集结方法研究[M]. 北京:经济科学出版社,2004
3. DeSanctis G. & Gallupe RB. A foundation for the study of group decision support systems [J]. Management Science, 1987, 33(5):589～609
4. Marcus Credle, Janet A Sniezek. Group judgment processes and outcomes in video－conferencing versus face-to-face groups[J],Int. J. Human-Computer Studies, 2003, (59): 875～897
5. 徐泽水. 不确定多属性决策方法与应用[M]. 北京:清华大学出版社,2004
6. Romero C, Tamiz M, Jones DF. Goal programming, compromise programming and reference point method formulations: linkages and utility interpretations [J]. Journal of the Operational Research Society, 1998,(49): 986～991
7. 斯蒂芬・P. 罗宾斯. 管理学[M]. 北京:中国人民大学出版社, 1998. 127
8. 岳超源. 决策理论与方法[M]. 北京:科学出版社,2003. 164
9. Mousseau V, Slowinski R. Inferring an ELECTRE TRI model from assignment examples [J]. Journal of Global Optimization, 1998, (12): 157～174
10. 唐・埃思里奇. 应用经济学研究方法论[M]. 北京:经济科学出版社,1998
11. Zadeh LA. Fuzzy Sets [J]. Information and Control. 1965
12. 温继文. 基于知识的鱼病诊断推理系统研究[D]. 中国农业大学,2003
13. 郭永洪. 基于本体的鱼病知识获取与诊断推理集成系统研究[D]. 中国农业大学,2004
14. Steve Cooke, Nigel Slack,制定管理决策教程[M]. 北京:华夏出版社,2000
15. George M. Marakas(朱岩,肖勇波译). 21 世纪的决策支持系统[M]. 北京:清华大学出版社,2002

方法篇

第 5 章

农业病虫害知识表示模型

知识表示是专家系统中十分重要的研究领域，与知识获取、推理、搜索、解释等有直接关系，对于问题能否求解以及问题求解的效率有重大的影响。本章在对农业病虫害诊断知识进行分类的基础上，分别研究案例知识、经验知识的知识表示和推理方式，并以鱼病和牛病为例，阐述了诊断知识的表示方法。

5.1 知识表示概述

知识表示水平的高低是决定农业病虫害诊断系统性能优劣的一个重要因素。合适的知识表示方法不仅可以表达出农业病虫害诊断领域的大量事实规律和专家经验，还对农业病虫害问题能否求解，以及问题求解的效率有重大的影响。因此，为了准确、细腻、真实地模拟农业病虫害专家的知识，有效求解农业病虫害诊断问题，研制开发一个经济、实用的农业病虫害诊断专家系统，首先要考虑的就是要选用合适的知识表示策略。

所谓知识表示就是将现实世界的知识转化为一种计算机可以接受的数据结构。由于计算机所能处理的知识需要满足一定的条件：①需要有统一的结构模式；②需要有限一致的符号；③由上述符号与模式构成合理的体系。自然语言作为一种描述现实世界的知识表示形式，并不符合上述条件，而且其多义性和模糊性也不利于计算机的处理。因此必须在现实世界与计算机世界之间建立“一方面比较符合人类较自然地表达其知识，另一方面又能变换成机器便于利用的格式”的中间变换界面，也就是我们所说的知识表示。[1] 针对农业病虫害诊断领域来说，就是首先将现实世界的领域知识经过人脑的抽象，转化为农业病虫害诊断概念化模型，然后将概念化模型转化为计算机世界的知识表示模型(见图 5-1)。

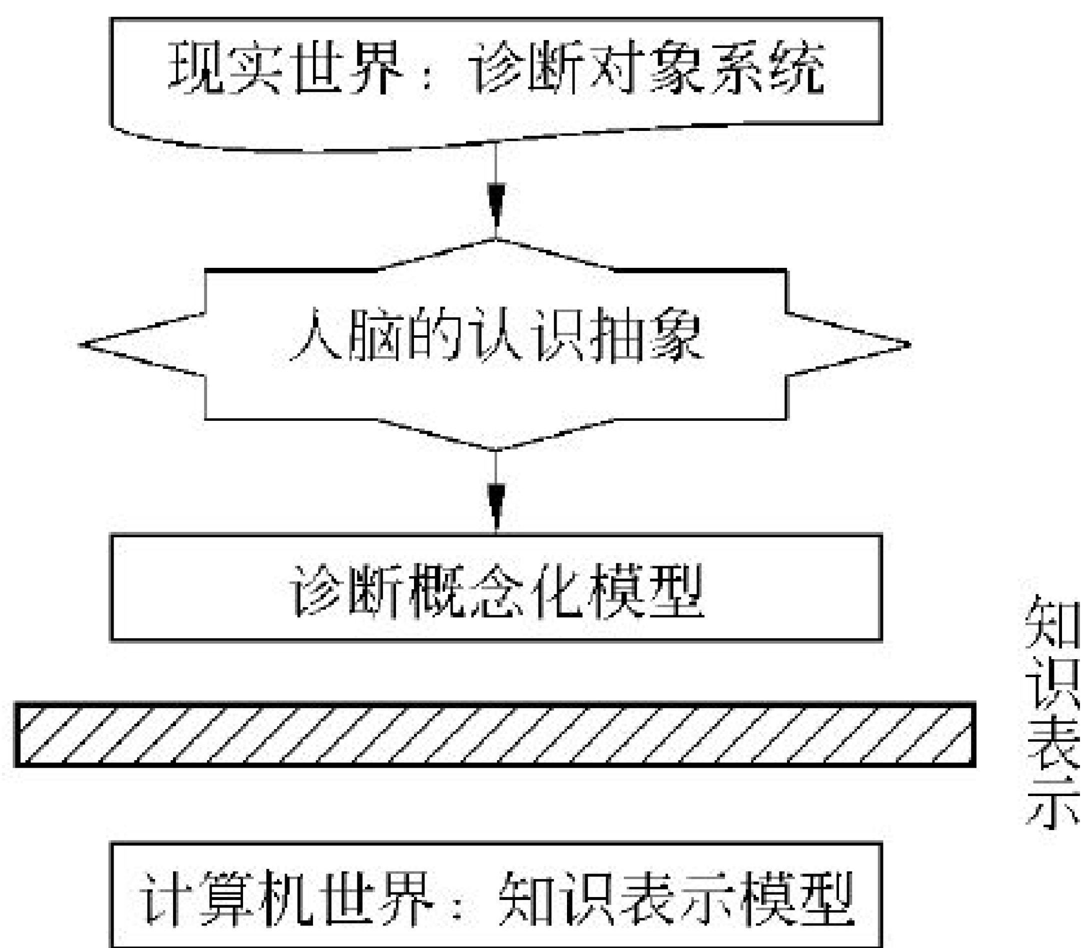


图 5-1 知识表示过程

目前，人工智能领域已经形成了多种知识表示方式，不同的知识表示方式效果不同，所以必须综合研究计算机知识表示的特点和农业病虫害诊断知识的特点，找出适合该领域知识特点和问题求解特性需要的最佳知识表示方式。

严格地说，知识表示是建立在知识分类基础之上的合理的形式化描述，因此我们在对农业病虫害诊断知识分类的基础上研究知识表示。

5.2 农业病虫害诊断知识的分类

一个基于知识的农业病虫害诊断系统至少应包括哪些类型的知识才有可能求解相应的问题？从不同的角度分析，知识具有不同的分类方法。

5.2.1 知识的作用

根据知识在智能程序求解问题过程中的作用，通常可分成四类：事实知识、规则知识、控制知识和元知识。[1]

(1) 事实知识：用来描述事实和概念，是有关问题环境的一些事实的知识，常以“……是……”的形式出现，如“肺炎是呼吸道疾病”。

(2) 规则知识：是有关问题中与事物的行动、动作相关联的因果关系知识，是动态的，常以“如果……那么……”的形式出现，例如“如果‘听诊时胸部有小片浊音区(多在肺脏肩前)或过清音’，那么就患了‘支气管肺炎’”。

(3) 控制知识：有关问题的求解步骤，技巧性知识，说明怎么做一件事。也包括当多个动作同时被激活时应选用哪个动作来执行的知识。

(4) 元知识：是有关知识的知识，是知识库中的高层知识。包括如何使用规则、解释规则、校验规则、解释程序结构等。元知识有时与控制知识是有重叠的。

5.2.2 知识的数据结构

按照数据结构形式，病虫害诊断知识又可以分为结构性知识、层次性知识和对象知识。[2]

(1) 结构性知识：以一定的结构存储在数据库中，如案例知识。规则知识在这里被认为是半结构性的，案例是规则的子集。

(2) 层次性知识：有一定继承关系的知识，如疾病的分类知识，如图 5-2 所示。

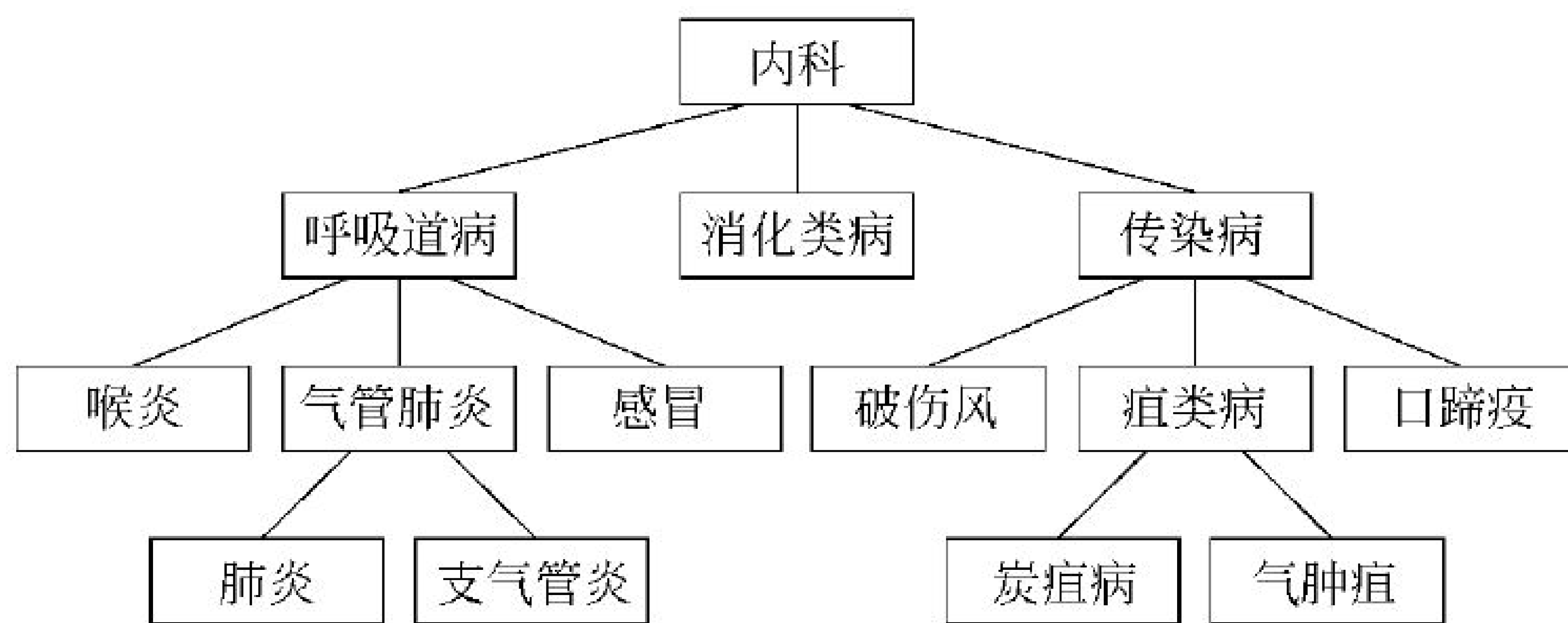


图 5-2 奶牛疾病层次分类图

(3) 对象知识：各种知识的综合体，它的属性包含各种症状、可能用到的规则等，它的方法包括各种计算规则、先验概率以及表示与案例相似程度的公式。

5.2.3 知识的确定性

若就知识的确定性来划分，知识可分为确定性知识和不确定性知识。确定性知识是指可指出其真值为“真”或“假”的知识。不确定性知识是指具有“不确定”特性的知识，它是对不精确、不完全及模糊性知识的总称。[1]



客观世界存在多种不确定性，大致可以分为三类：随机性、模糊性和不完备性。随机性是指由于事件之间的因果关系不明确，因而不能完全确定事件发生结果的一种不确定性；模糊性是指人们对现象或知识的概念外延不清晰而导致的一种不确定性；不完备性是指由于人们的认识和客观条件的限制，特征信息不完全或不精确而导致的一种不确定性。专家系统的一个重要任务就是要构造一个能够合理处理这些不确定性知识表示的模型，其前提是需要考虑模型的有效性和实用性。所谓有效性是指理论基础扎实，对模型限制少，应用简便而且效果好；实用性是指证据和知识的不确定性容易描述，度量值容易给出。[2] 主要有以下几种表示法。

(1) 经典关系表示法：用一个取值集合为{0,1}的经典二元关系表示事件之间的不确定性。典型例子为J. A. Keggia于1985年首次提出的覆盖集模型GSC(model based on general set-covering)模型，该模型用一个经典二元关系来刻画事件之间的因果关系，在此基础上提出一些基本概念，并讨论了问题解的求法。

(2) 概率关系表示法：用一个取值范围为[0,1]的概率关系表示事件之间的不确定性。贝叶斯方法是基于概率关系表示的一种方法，知识完全以事件发生的先验概率以及事件之间的条件概率表示，概率的获取来自于大量数据的统计或由领域专家提供，当领域问题较为复杂或涉及大量问题时，概率的获得变得非常困难。

1975年肖特里菲(Shortliffe)和Buchanan基于确定性理论并结合概率论提出了可信度方法，并将其成功应用于MYCIN系统；杜达(R. O. Duda)和哈特(P. E. Hart)于1976年提出了主观贝叶斯方法，并在地矿勘探专家系统PROSPECTOR中得到了成功的应用。由于主观贝叶斯方法也存在概率值获取困难的问题，Dempster提出了证据理论，Shafer作了进一步的完善，所以称为D-S模型。证据理论不要求专家被迫给出每个事件的先验概率，只要求给出基本概率的分配函数，并且具有利用证据的积累，缩小假设集合的能力。[4]

(3) 模糊关系表示法：用一个取值区间为(0,1)的模糊关系表示事件之间的不确定性。

基于概率的推理虽然可以表示和处理现实世界中存在的不确定性，但不能反映事物本身的模糊性，1978年扎德等人提出的模糊集理论以及模糊逻辑弥补了这一缺憾，开辟了基于模糊信息的推理方法。模糊方法认为事件之间的关系是Fuzzy关系，具有形式：$Y=X\cdot R$。

5.2.4　知识的表现形式

按照知识的表现形式可将知识分为：逻辑性知识和形象性知识。逻辑性知识是反映人类逻辑思维过程的知识，例如人类的经验性知识等。这种知识一般都具有因果关系及难以精确描述的特点，它们通常基于专家的经验以及对一些事物的直观感觉。人类的思维过程除了使用逻辑思维方式外，还有一种称为“形象思维”的思维方式。通过事物的形象建立起来的知识称为形象性知识。[5]

5.3　病虫害知识的表示

5.3.1　案例知识的表示

案例是对事件或问题的描述和表示，其内容包括问题描述、问题求解和案例结果三部分。本节以鱼病案例为例，来讲述案例知识的表示模型。

鱼病案例包括三部分，即鱼病诊断的问题描述、鱼病诊断的问题求解和鱼病诊断的

结果。[6]

(1) 鱼病诊断的问题描述：鱼病诊断的问题描述是描述鱼病案例发生时的背景和状态，是对案例表示问题的各种属性特征的描述，是鱼病发生的自然情况和症状描述，包括鱼的品种、规格、发病时间和症状表现等。

(2) 鱼病诊断的问题求解：鱼病诊断的问题求解是对问题描述选择处理方法和解决策略，得到相关结论、解释或建议，即根据病鱼的症状描述推理得出确诊的疾病、病因和防治措施。

(3) 鱼病诊断的结果：鱼病诊断的结果是指事件发生后的状态，即此案例是成功案例还是失败案例等，是诊断的效果反馈。

在鱼病诊断推理系统中，根据鱼病诊断的特征和内容，以特征向量作为案例的表示形式。一个新的目标案例 case 含有信息即症状集 $S^*=\{s_1,s_2,\cdots,s_x\}$，则从案例库中开始搜索，这里 $S^*\subseteq S=\{s_1,s_2,\cdots,s_n\}$，按照搜索算法，从 $DC=f(DP)=f(DO,D,S,P,C,R)$ 中以 S^* 为检索关键字，检索出 $case^*=case=\{DO,D,S^*,P,K,R\}$，则为诊断结果，将其输出。

传统的鱼病诊断记录是将各种鱼病的诊断内容记录到相同结构的数据表或数据库中，字段如下：(病历编号，日期，种类，规格，溶氧值，池中表现，头部症状，鳃部症状，体表症状，内脏症状，肌肉症状，鳞片症状，目检结论，镜检水体，镜检鱼体，病因，治疗方案，医生姓名)。采用相同的数据结构记录病例可使数据库结构简单，但也有很多缺点，突出表现为数据库冗余量大。因为这一结构包括了病鱼可能出现的所有情况的属性，但是每个案例只表现出其中的一部分属性，这样就导致很多空值，造成数据库资源的浪费。

通过对鱼病案例的分析发现，不同种类的鱼病所表现出来的症状特征属性既相互区别，又具有聚类的特点，即某种或某些鱼病仅在特定部位才有症状表现，其他部位则基本不显示症状。这样就为鱼病案例的知识表示提供了进一步细分的可能，如表 5-1 所示。

表 5-1　不同鱼病的症状在不同部位的表现

病历编号	池中表现	头部症状	鳃部症状	体表症状	内脏症状	肌肉症状	鳞片症状	疾病
1	1	0	1	1	1	0	0	肠炎
2	1	0	0	0	1	0	0	肠炎
3	1	0	0	1	1	0	0	肠炎
4	1	0	0	1	1	0	0	肠炎
5	1	1	1	0	0	0	0	烂鳃
6	1	0	1	0	0	0	0	烂鳃
7	1	0	1	0	0	0	0	烂鳃
8	1	0	1	1	0	1	0	烂鳃
9	1	1	1	1	1	0	1	竖鳞病
10	1	1	1	1	1	0	1	竖鳞病
11	1	0	0	1	1	0	1	竖鳞病
12	0	1	1	1	1	0	1	竖鳞病
13	1	0	1	0	0	0	0	寄生虫
14	1	0	1	0	1	0	0	寄生虫
15	1	0	1	1	0	0	0	寄生虫
16	1	1	0	0	0	0	0	寄生虫

注：“1”表示该部位有症状；“0”表示该部位无症状。



以肠炎病为例，病鱼的症状表现主要发生在池中、鳃部、体表、内脏，如表 5-2 所示。按照以往的症状记录形式，每个部位对应着一系列症状描述，分别用逗号隔开，推理时对这些症状单元逐一进行检索。

表 5-2　肠炎病症状表现

案例编号	鱼的池中表现	鳃部症状	体表症状	内脏症状
1	几乎不摄食，逐渐有死亡个体	略有贫血或红紫	腹部有的肿胀，肛门红肿，有黄色液体流出	肠道充血，无食物，有脓液，部分有腹水
2	少数在池边打转，连续几天池边有死鱼	正常	正常	肠道积水，或肠壁充血，肠道无食
3	不吃食，死亡	正常	肛门红肿外突	肠道积水有脓液
4	厌食，少见死亡	黏液多	肛门红肿外突	肠内充血积液，无食物
5	吃食鱼渐少，死鱼渐多，离群独游	有黏液增多现象	发黑，腹部膨大，肛门红肿外突	肠壁充血，内无食物，黏液多，有时有脓
6	吃食不好	正常	正常	肠道发红
7	吃食不好	鳃丝充血，分叉	正常	肠道发红，肠内有剩料
8	吃食不好，游动缓慢	正常	发黑，腹部肿胀，有的有红斑，肛门红肿，轻压腹部有乳黄色液体流出	肠道发红
9	吃食不好	正常	正常	肠道发红

考虑从表 5-2 按部位区分症状的方法转换为按照单个症状单元的记录方式，就是从每个部位的一系列症状表现中抽取出互异的症状单元，如表 5-3～表 5-5 所示。

表 5-3　肠炎病症状的符号表示

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
吃食不好	不吃食	有死亡	腹部肿胀	肛门红肿	肛门外突	肛门有黄色液体流出	肠道充血	肠道积水	肠道无食	肠道有脓液

表 5-4　肠炎病的症状符号表示

案例编号	鱼的池中表现	体表症状	内脏症状
1	C2，C3	C4，C5，C7	C8，C9，C11
2	C3	0	C8，C9，C10
3	C2，C3	C5，C6	C9，C11
4	C1	C5，C6	C8，C9
5	C2，C3	C4，C5，C6	C8，C10，C11
6	C1，C3	0	C8
7	C1	0	C8
8	C1	C4，C5，C7	C8
9	C1	0	C8

表 5-5　肠炎病症状的 0-1 表示

案例编号	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
1	0	1	1	1	1	0	1	1	1	0	1
2	0	0	1	0	0	0	0	1	1	1	0
3	0	1	1	0	1	1	0	0	1	0	1
4	1	0	0	0	1	1	0	1	1	0	0
5	0	1	1	1	1	1	0	1	0	1	1
6	1	0	1	0	0	0	0	1	0	0	0
7	1	0	0	0	0	0	0	1	0	0	0
8	1	0	0	1	1	0	1	1	0	0	0
9	1	0	0	0	0	0	0	1	0	0	0

从表 5-2 可以看出，某种鱼病的症状总是集中在某几个部位，而其他部位则不显示症状或很少出现症状。因此，将鱼病案例库按照不同鱼病划分为多个子案例库，对于每种鱼病，在症状表现的属性中保留那些症状出现频繁的部位，忽略那些不显示症状或很少出现症状的部位，使得数据库冗余减少。因此，鱼病诊断的案例知识表示为由二元组转换为 $(n+1)$ 元组，其中 $n=$ 该种鱼病的症状属性个数：

$$\text{case}_n=(s_1,s_2,\cdots,s_{n-1},s_n),\quad s_i=\begin{cases}0, & \text{没有症状 } s_i\\ 1, & \text{有症状 } s_i\end{cases},\quad i=1,2,\cdots,n$$

5.3.2　经验知识的表示

根据专家系统中一般知识的分类方法，经验知识可分三个级别：概念级、事实级、规则级。[7] 本节以鱼病诊断知识为例介绍经验知识的表示方法。

(1) 概念知识：对应于数据库中的属性值，如症状、疾病、病因等概念和定义等。

(2) 事实知识：建立概念间的联系，对应于数据库中的元组。以环境、症状、病因和治疗知识为例，相应的知识表示模式如下。

① 环境知识：是对鱼发病过程中水环境理化指标的一种定性或定量的描述，通常情况下可作为鱼病专家进行疾病诊断和病因分析的依据。

(〈水体环境知识框架〉∷=〈水体的溶氧度〉|〈水体的氨氮值〉|〈水体的 pH 值〉
〈池塘深度〉|〈水中的藻类〉)

② 症状知识：是对鱼体表现异常的定性和定量的描述。通常症状是鱼病专家进行疾病诊断的直接依据。

(〈症状框架〉∷=〈症状代号〉|〈症状名称〉|〈症状类型〉)

〈症状类型〉∷=症状的重要性，取值为 1 表示主要症状，取值为 0 表示辅助症状。

③ 病因知识：是对鱼体原因的一种定性描述，通常病因知识分为生物致病因素(细菌、真菌、病毒、寄生虫)、环境致病因素(水温、溶氧、pH 值等)和人为因素(机械损伤、饲料管理等)。

(〈病因框架〉∷=〈病因代号〉|〈病因名称〉|〈病因类型〉)

④ 治疗知识：指鱼体发生异常或疾病时是否应该采取或采取怎样的治疗措施的知识，病因知识必须在综合分析鱼病和病因基础之上得出。

(〈治疗知识框架〉∷=〈鱼病名称〉|〈鱼病原因〉|〈治疗方法〉)

(3) 规则知识：主要指上述四类事实知识之间的关系，对应于专家系统知识库中的规则



知识。例如，竖鳞病的诊断规则表示如下：

IF　现场表现：游动缓慢，呼吸困难，身体倒转，腹部向上，有死亡
　　AND 体表：鱼体呈松球状，皮肤充血
　　AND 鱼鳍：鳍基充血
　　AND 头部：眼球突出
　　AND 腹部：腹部膨胀，腹腔内积水
　　AND 鳞片：鳞的基部水肿，鳞片竖起，鳞片脱落，脱鳞处形成红色溃疡

THEN　鱼病：竖鳞病
　　AND 预防：鱼体受伤是此病的主要原因之一，因此，在牵捕、搬运、放养等操作过程中，应注意防止鱼体受伤
　　　　AND 鲤鱼产卵池在冬季要进行干池清整，并用生石灰或漂白粉消毒
　　AND 治疗：用链霉素或氯霉素进行腹腔注射，每尾用药3～6 mg
　　　　OR 每100 kg水加捣烂的大蒜0.5 kg，搅匀后给病鱼浸洗数次
　　　　OR 用2%食盐与3%小苏打混合液给病鱼浸洗10分钟
　　　　OR 3%食盐水浸洗病鱼10～15 min
　　　　OR 遍撒消毒剂，严重时隔日再用一遍
　　　　OR 内服抗菌素
　　　　OR 全泼二氧化氯
　　　　OR 用4‰～5‰的新诺明作药饵连用3～5天，或外用杀菌药，内服土霉素药料5～7天
　　　　OR 定期泼洒生石灰
　　　　OR 全池泼洒漂白精

5.4　小结

本章在对农业病虫害诊断知识进行分类的基础上，分别研究案例知识、经验知识的知识表示和推理方式。经验知识又分为概念知识、事实知识和规则知识，以鱼病诊断知识为例阐述了其表示方法。

参考文献

1. 王永庆. 人工智能原理与方法[M]. 西安：西安交通大学出版社，1998
2. 高济. 基于知识的软件智能化技术[M]. 杭州：浙江大学出版社，2000.9
3. 温继文. 基于知识的鱼病诊断推理系统研究[D]. 中国农业大学，2003
4. 郭永洪. 基于本体的鱼病知识获取与诊断推理集成系统研究[D]. 中国农业大学，2004
5. 钟东屏. 诊断逻辑学[M]. 贵州：贵州科技出版社，1991
6. 周云. 基于案例推理的鱼病诊断专家系统研究[D]. 中国农业大学，2004
7. 郑育红. 网络化鱼病诊断专家系统[D]. 中国农业大学，2000

第 6 章

基于遗传算法的鱼病案例推理模型

案例学习是基于案例的推理(CBR)中的关键技术问题之一,本章提出一种利用遗传算法来完成特征项权空间的直接搜索的算法,根据鱼病诊断的特征及内容,对基于案例的推理方法进行了系统研究,并将其应用于鱼病诊断系统中。

6.1 基于案例推理的基本概念

CBR 是一种问题求解方法,与其他推理方法相比有着本质的不同。它能通过对比,从过去同类问题的求解方法和经验中寻找出适合当前问题的解决方案,并在当前问题的求解过程中不断对过去的经验和知识进行修正、学习,记录新的经验和知识,用于解决后来出现的问题。[1]

人们常采用基于案例的推理来提高解决问题的效率。例如,当一个鱼病医生面对具有多种症状的病鱼时,如果曾见过具有类似症状的病鱼,医生往往回忆起旧的病例并参考旧病例的诊断,从而对新病鱼作出诊断。如果每一次进行鱼病诊断都不参考旧的病例,而是完全基于鱼病诊断知识和测试手段,通常需要花费很多时间,而且可能得不出结论,这一点可以从没有经验的鱼病医生那里得到实证,由此可以看出利用过去的经验可以使鱼病医生节省大量诊断时间。

6.1.1 基于案例推理的基本原理

CBR 是近十几年来人工智能中发展起来的区别于基于规则推理的一种推理模式,它利用旧的案例或经验来解决问题、评价解决方案、解释异常情况或理解新情况。案例是一段带有上下文信息的知识,该知识表达了推理机在达到其目标的过程中能起关键作用的经验。在 CBR 中,把当前面临的问题或情况称为目标案例,而把记忆的问题或情况称为源案例。[2]

CBR 的形成及其发展有着很强的理论背景,主要体现在以下两个方面。

(1) CBR 的心理学模型

CBR 是人的一种认知行为,本质上是基于记忆的推理。随着记忆认知理论的发展,心理学家提出了语义网络记忆模型以描述事物及其内部联系。由于这一模型只表示了关于世界的静态知识,Tulving 提出了情景记忆的概念,即人对一定时间内所发生事件的记忆,并将情景记忆作为语义记忆的补充;与此同时,以 Schank 为代表的人工智能(AI)研究者在自然语言理解的研究中发展了类似理论,提出了概念记忆(conceptual memory),认为记忆本质上是情景化的,信息来源于经历,知识不是固有的,记忆理论必须解释知识的获取。

情景记忆、概念记忆的提出奠定了 CBR 的心理学基础,它包含以下心理学假设:①我们记



忆了一个个的案例(case);②世界是有规律的,类似的问题有类似的解决方案,因此,以前类似问题的解决方案可以作为解决新问题时有用的起点;③遇到的问题类型还会重现,因此,未来的问题可能与现在的问题类似。当这几个原则都成立时,CBR 就是一种有效的推理策略。而且,CBR 对于解决与以前经验完全不同的问题也有帮助。当基于案例的推理将案例应用到异常问题时,可以通过简单重用或者更有创造性的问题求解过程来处理变化。这几点本质上意味着:记忆是情景化的,知识来源于经历。

(2) CBR 与模拟思维

案例推理是人类直觉、逻辑和创造性三种思维的一种综合表现形式。案例推理符合人类的思维活动。遇到新事物时,专家不仅仅看到一个具体问题,还会产生联想,然后把事物归类,从中找出以往处理类似问题的经验和相关知识,经过一定的修正处理后用以解决新的问题。对于简单的问题,案例的检索和匹配主要是形象思维的过程。而对于较为复杂的问题,一般难以通过简单匹配检索到一个相似的案例;这时人们往往会分解问题,使每个子问题能映射到一个相似案例,或从不同的角度出发,抽取出不同角度的类似问题,最后运用逻辑思维和创造性思维把匹配的子案例集成起来,形成解决当前问题的新方法。

CBR 中,案例库模拟人脑的记忆,存储了一些过去的相关经历(案例),这些案例按一定的方式组织起来,以便在需要时及时取出,即 CBR 中从案例库中检索出相关案例的过程。检索出的候选案例可能与新情形不完全一致,这时需要对候选案例的某些特征进行修改,以适应新的情况。修正后的案例是否适合实际情况,还需检验。如果经过检验发现不符合实际情况,还要修正,最终新的案例以一定的策略加入案例库中。从上述过程可以看出,CBR 技术之所以成为研究的热点,不仅在于其推理技术相对简单,而且在于它完全区别于 RBR(基于规则的推理)、MBR(基于模型的推理)。这是因为专家解决问题往往不是通过抽象的知识规则去演算,而是首先基于对自身经历和经验的回忆。与借助于已有案例来解决新问题一样,通过对旧案例的证实和修正,可不断获得新的知识,因此,应用 CBR 技术,可以达到动态知识学习的目的。[3]

6.1.2　基于案例推理的一般过程

CBR 任务通常可以分为两类:解释型 CBR 和问题求解型 CBR。前者将源案例作为划分新案例的参考点;后者用源案例来提供适用于新案例的解决方案。解释型 CBR 的目标是将新案例与已经分类的案例进行比较,形成对新案例的判断或分类。它常用于概念解释、诊断等领域。而问题求解型 CBR 的目标是运用以前的解决方案来产生新问题的解决方案。它常用于设计、计划、解释等领域。[4]显然,鱼病诊断系统的推理机应该运用解释型 CBR,目的是为了根据已有的鱼病诊断源案例的推理方式,得出现有的目标案例的结论,从而找到治疗新鱼病的方案。

CBR 实际上是一个解决问题、从经验中学习、再解决一个新问题的循环和集成过程,包括如下四个子过程:案例检索、案例重用、案例修改和案例学习。案例检索是指从案例库中检索与新问题最相似的案例;案例重用是指重用检索出的案例的信息和知识来解决新问题;案例修改是指修改求解方案;案例学习是指学习新的经验并把它存入现存的案例库中(见图 6-1)。

根据以上描述,可以发现 CBR 的核心问题包括五个方面的内容:案例表示方法、案例索引、案例检索、案例修改和案例学习。

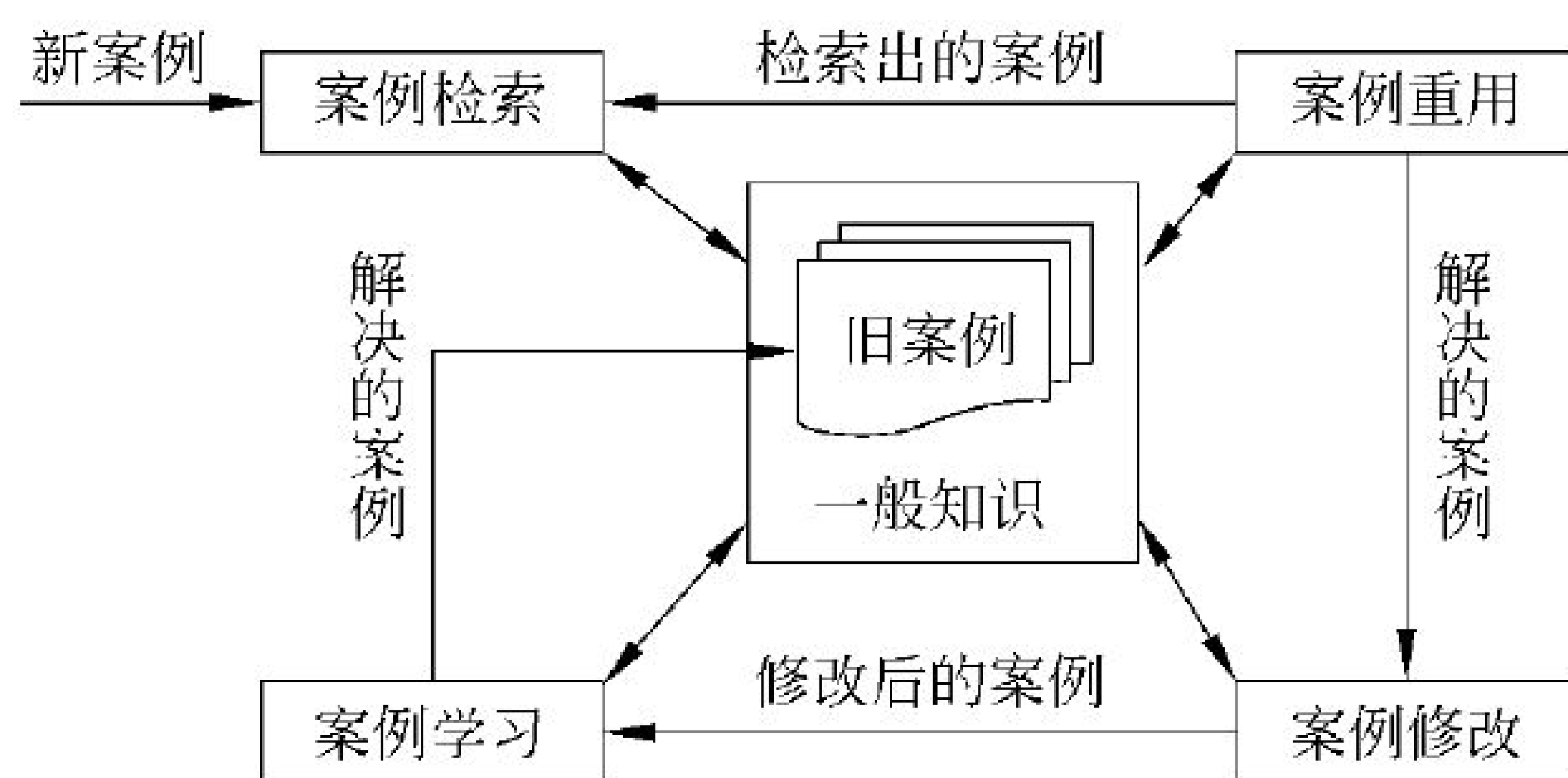


图 6-1　CBR 循环周期

(1) 案例表示

案例表示的研究，是整个 CBR 研究的基础和核心。就简单的形式而言，案例就是能导致特定结果的一系列特征属性的集合，如鱼体某部位的症状及其诊断结果就是鱼病诊断的一个案例；就复杂的形式而言，案例就是形成问题求解结构的子案例的关联集合，如某一并发症由该并发症的子疾病的症状组成，每个组成部分对整个问题来说可以被认为是一个案例。鱼病诊断推理系统中案例表示的主要任务是确定适当的案例特征，包括鱼病诊断领域专有名词的定义和专家诊断过的鱼病典型案例收集。

(2) 案例索引

CBR 系统的强大功能在于它能从其记忆库中迅速准确地检索出相关案例。案例索引的目标是在相似的情形下，应该检索出什么样的案例。通常有三种案例索引技术：最近邻法、归纳法(决策树法)和知识索引法。也有采用这三种方法结合的方法。

最近邻法是指用户利用与记忆库案例相匹配的输入案例的特征权数和来检索案例。单独使用该方法的最大问题是不可能覆盖全局特征的权值集，有些情况下难以准确检索到案例。许多问题的特征权重是相互依赖的，即在确定要检索的适当案例时，一个给定特征的重要性取决于这个案例其他特征的权值。

归纳法有两个优点：一是能自动、客观、严格地分析案例，能区别这些案例的最佳特征；二是案例可以组织成分层结构，其检索时间呈对数增加。主要缺点是：当使用的案例库包括数以千计的案例时，为了完成归纳，系统需要相当数量的案例来生成判别树，归纳分析时间长。

知识的索引法利用现存的案例相关知识，在检索案例时用以确定哪些特征是重要的，缺点是常常难以代码化解释性知识，难以在大范围的案例库中完成完备的知识索引。因此，许多系统使用与其他索引技术相结合的基于知识的索引法。

(3) 案例检索

一旦完成了案例表示和索引，一个高效的检索结构就组织好了。大多数案例存储结构介于纯粹的关联检索与纯粹的分层检索之间。前者指一个案例的一些或全部特征与其他特征相互独立地索引；后者指案例特征被组织成一般到特殊的概念结构。Schank 的动态记忆模型和 Porter 的分类案例模型是比较有影响的两种存储组织模型，动态记忆模型呈一种层次结构，分类案例模型是一种网状结构，由分类语义关系、案例及索引指针构成。

(4) 案例修改

案例修改是指将一个检索到的与当前情况大部分匹配的案例修改成完全匹配的案例。现有 CBR 系统的大多数修改方法是将修改知识代码化成修改规则集或领域模型。修改规则可



使检索到的案例转化为一个满足所有要求的新案例。许多CBR系统是利用所存储的现有案例来完成修改的，而且即使案例不能被计算机修改，至少也为人工修改提供了一个有意义的起点。

（5）案例学习

案例学习对CBR系统很重要。利用现有技术从案例库中抽取有用信息，即收集问题求解知识并正确利用。随着案例的积累，案例学习可以确定体现特定案例群主要特征的原型案例，并且原型案例可与特定案例存储在一起，提高系统长期运行的准确性。

6.2　鱼病诊断推理系统中的案例表示

CBR有两个基本组成部分，即案例和案例推理。其中案例是一种结构化的知识表达方式，是CBR技术中操作和处理的主要对象，它是专家求解问题的实际例子，包含了专家处理问题时所处的环境信息、所应用的领域知识、所采取的求解方案以及方案的实施结果等内容，所以表示一个案例的关键知识必须完备，选择的特征属性要合适，不可忽视具有一定影响的特征属性，不可加入对原问题无关的特征属性。

案例表示具有两方面的含义，即内容与结构。它是指用模型化处理的方式在计算机内部反映案例的数据及其关系，以便于实现对案例的操作。案例的表示方法不仅反映案例能否方便地转化为案例的形式，而且还影响着案例推理的效率及求解精度。

6.2.1　案例的主要内容

一个案例是对一个事件或问题的描述和表示，案例表示的内容包含了与这些事件或问题相关的各种不同信息。案例通常包括下述三个主要组成部分。

（1）问题描述

问题描述是描述案例发生时的条件，即事件发生时的背景和状态，是对案例表示问题的各种属性特征的描述，一般与问题相关的特征、属性以及这些特征属性在问题中所起作用的大小有关，主要指案例中所描述问题出现或发生时的系统特征、环境状态、场景属性，以及其他与问题处理相关的各种因素。

（2）问题求解

问题求解是对问题的求解过程或求解方法的描述，即对案例所描述的问题选择处理方法和解决策略，得到相关结论、解释或建议。对问题处理过程的描述可以根据不同情况和需要采用不同的方式，有时只需简单地记录问题处理方法和解决策略本身的结果或结论，有时则需要对推导或应用过程进行描述。

（3）案例结果

案例结果是指事件发生后的状态，是成功案例还是失败案例等情景。

问题描述和环境描述是案例的主要内容，是进行推理的基础，也是案例建立并需要研究的重要问题，对求解问题有着重大影响。

6.2.2　案例的表示方法

知识在人脑中的记忆本质至今依然无法获知，因此案例的表示也一直缺乏统一的方法，CBR的主要创始人Schank提出用脚本来表示案例，脚本主要是用来记录与某一特定情景相

关联的特定思想,通常用一系列特征来描述案例。根据这一思想,目前知识系统中采用了产生式、框架、特征属性、谓词、数据库、语义网、面向对象的结构化表示等方法。

外部表示与内部表示是案例表示的两种表现形式。外部表示是人机交互意义下的案例表示,强调表示手段的自然以及符合人类的思维习惯,而内部表示则是指案例在计算机内部的表示,强调要适合计算机处理。

除此之外,还应当注意到这样一个问题:当案例描述一个复杂情景时,将复杂的情况分解成不同层次上的简单情况,是人们通常会采用的手段,这种分解可以从抽象特征和具体特征上来考虑。抽象层次是指案例描述中所描述的特征的具体程度。采用易于感知的特征称为具体表示,使用抽象特征称为抽象表示。具体表示内容详尽、信息量大,但不易于管理和操作,特别是对检索操作,而抽象表示形式简单,便于高效检索,也便于匹配和修改,但抽象特征提取困难,信息易丢失且易以偏概全。一个案例中,通常既包括抽象特征又包括具体特征,在鱼病诊断推理系统中对于抽象层次使用基于规则的产生式方法表示,而对于具体特征则采用具体案例来表示。

常用的案例表示方法是用一组特征变量来记录一个案例的所有特征,利用案例的特征反映案例内容,这种方法实现起来比较简单。通常,一个典型案例至少应包括一个问题描述部分和一个解决方案描述部分,这可以用二元组〈问题,解决方案〉来表示。对于鱼病诊断系统来说,问题就是鱼所表现出来的症状、水质参数等;解决方案就是找出疾病及成因,并给出相应治疗方案。

6.2.3 鱼病诊断推理系统案例表示

在鱼病诊断推理系统中,根据鱼病诊断的特征和内容,以特征向量作为案例的表示形式。为了叙述方便,我们以目检过程为例,省略了具体案例中与案例推理关系不大的疾病现象与防治方法等内容,只关注与鱼病有关的案例的特性属性。

以鲤鱼为例,在发生疾病时表现为肌肉、体表、腹部、鳞片、头部、鱼鳍、鳃部、肠道等异常症状。因此,在对该病鱼进行诊断时,就用上述 8 个参数作为特征来建立其疾病案例。表 6-1 中列举出根据这些特征所建立的鲤鱼发生疾病时的 5 个案例,每个案例代表其在疾病状态下所表现出的不同的特征组合。在这些案例中没有考虑特征权重对案例的影响,即假定各个参数的作用和影响是同等的。

表 6-1 鲤鱼疾病案例

ID	疾病	肌肉	体表	腹部	鳞片	头部	鱼鳍	鳃部	肠道
Case1	白头白嘴病	000	A01	000	000	B04	000	000	000
Case2	肠炎病	000	000	D01	000	000	000	000	H03
Case3	赤皮病	000	A04	000	E01	000	F01	C03	000
Case4	烂鳃病	000	000	000	000	B01	000	C01	000
Case5	疖疮病	000	A07	000	000	000	F02	000	000

如果把每一个案例作为一个行向量,则可用以下矩阵 $\boldsymbol{C}$ 表示上述 5 个案例集。



$$
\boldsymbol{C}=\begin{bmatrix} 000 & A01 & 000 & 000 & B04 & 000 & 000 & 000 \\ 000 & 000 & D01 & 000 & 000 & 000 & 000 & H03 \\ 000 & A04 & 000 & E01 & 000 & F01 & C03 & 000 \\ 000 & 000 & 000 & 000 & B01 & 000 & C01 & 000 \\ 000 & A07 & 000 & 000 & 000 & F02 & 000 & 000 \end{bmatrix}
$$

在实际应用中，不同特征对案例的作用和影响存在着一定的差别，只是这种差别在不同情况下的表现程度不同而已。针对表6-1的案例，若考虑8个特征对案例的影响和作用上的差别，对它们赋予不同的权值后，可以得到一个新的案例集合，如表6-2所示。

表6-2　新的鲤鱼疾病案例

ID	疾病	肌肉	体表	腹部	鳞片	头部	鱼鳍	鳃部	肠道
Case1	白头白嘴病	000	A01	000	000	B04	000	000	000
	权值	0	0.5	0	0	0.5	0	0	0
Case2	肠炎病	000	000	D01	000	000	000	000	H03
	权值	0	0	0	0	0	0	0	0.5
Case3	赤皮病	000	A04	000	E01	000	F01	C03	000
	权值	0	0.25	0	0.25	0	0	0.25	0
Case4	烂鳃病	000	000	000	000	B01	000	C01	000
	权值	0	0	0	0	0.5	0	0.5	0
Case5	疥疮病	000	A07	000	000	000	F02	000	000
	权值	0	0.5	0	0	0	0.5	0	0

由于在不同案例中，特征以及特征的权重都可能发生变化，如上例中烂鳃病的症状及权重表现在头部和鳃部，而肠炎病的症状及权重却表现在腹部与肠道，因此该例是一个非同构案例集合，为此采用案例集合的最大特征集来解决非同构问题。

可以根据最大特征集合来组织案例特征矩阵 $\boldsymbol{F}$：

$$
\boldsymbol{F}=\begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,m} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,m} \\ \vdots & \vdots & & \vdots \\ f_{n,1} & f_{n,2} & \cdots & f_{n,m} \end{bmatrix}
$$

在以上矩阵中，m 为最大特征集合中的特征个数，即最大特征数；n 为案例集中的案例个数。

在特征矩阵 $\boldsymbol{F}$ 中，每一行代表一个案例，每一列代表一个案例特征在不同案例中的特征值，对矩阵中的每一个元素 $f_{i,j}(1\leqslant i\leqslant n, 1\leqslant j\leqslant m)$ 的取值，作如下规定：

$$
f_{i,j}=\begin{cases} \text{特征 } j \text{ 在案例 } i \text{ 中的取值}, & \text{当特征 } j \text{ 是案例 } i \text{ 中的一个特征时} \\ \text{无意义}, & \text{当特征 } j \text{ 不是案例 } i \text{ 中的一个特征时} \end{cases}
$$

上式中当特征 j 是案例 i 中的一个特征时，它表示特征 j 在案例 i 中的取值，否则无实际意义，可用任意适当的数值来填充该元素在矩阵中的位置。

类似地，我们把案例中各特征的权重也按照与上述相似的方式组织成一个矩阵，并称该矩阵为特征权矩阵。

$$\boldsymbol{R}=\begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,m} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,m} \\ \vdots & \vdots & & \vdots \\ r_{n,1} & r_{n,2} & \cdots & r_{n,m} \end{bmatrix}$$

上式中,$\boldsymbol{R}$ 为特征权矩阵;m 为最大特征集合中的特征个数;n 为案例集合 C 中的案例个数。

在特征权矩阵 $\boldsymbol{R}$ 中,每一列代表一个案例特征在不同案例中的特征权值,对矩阵 $\boldsymbol{R}$ 中的每一个元素 $r_{i,j}(1\leqslant i\leqslant n,1\leqslant j\leqslant m)$的取值作如下规定:

$$r_{i,j}=\begin{cases}\text{特征 } j \text{ 在案例 } i \text{ 中的取值}, & \text{当特征 } j \text{ 是案例 } i \text{ 中的一个特征时} \\ 0, & \text{当特征 } j \text{ 不是案例 } i \text{ 中的一个特征时}\end{cases}$$

当特征 j 是案例 i 中的一个特征时,$r_{i,j}$表示特征 j 在案例 i 中的权值,否则 $r_{i,j}=0$。

在上述定义的基础上,一个案例集合 C 可表示为该案例集合的特征矩阵与特征权矩阵的转置矩阵的乘积:

$$\boldsymbol{C}=\boldsymbol{F}\cdot\boldsymbol{R}^{\mathrm{T}}$$

例如,我们利用矩阵可以把由表 6-2 所示的鱼病案例集合表示为

$$\begin{bmatrix} \boldsymbol{C}1 \\ \boldsymbol{C}2 \\ \boldsymbol{C}3 \\ \boldsymbol{C}4 \\ \boldsymbol{C}5 \end{bmatrix}=\begin{bmatrix} 000 & \mathrm{A}01 & 000 & 000 & \mathrm{B}04 & 000 & 000 & 000 \\ 000 & 000 & D01 & 000 & 000 & 000 & 000 & \mathrm{H}03 \\ 000 & \mathrm{A}04 & 000 & \mathrm{E}01 & 000 & \mathrm{F}01 & \mathrm{C}03 & 000 \\ 000 & 000 & 000 & 000 & \mathrm{B}01 & 000 & \mathrm{C}01 & 000 \\ 000 & \mathrm{A}07 & 000 & 000 & 000 & \mathrm{F}02 & 000 & 000 \end{bmatrix}$$

$$\cdot\begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0.25 & 0 & 0.25 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$$

6.3 鱼病诊断推理系统案例的检索

案例检索是从案例库中查找和搜索出与当前问题在属性特征上近似并对当前问题的求解具有帮助或指导意义的案例,直到找到一个最好的匹配。这个过程通常包括初级匹配、搜索和选择。初级匹配的目标是返回一系列案例,能与新案例充分一致。案例的检索过程通常分为两个步骤进行,即案例索引和案例匹配。案例索引就是按照一定的索引策略,先从众多案例中初步筛选出对当前问题可能或潜在有用的案例集,以缩小检索范围,提高查找速度;案例匹配就是从经过初选得到的案例集中根据一定的相似性度量方法查找出与当前问题最相似的案例。[5]

6.3.1 现有的案例检索模型

CBR 的检索要达到以下两个目标:一是检索出的案例尽可能的少;二是检索出来的案例尽可能与目标案例相关或相似。检索和选择的关键是进行案例相似性的度量。检索策略有很多种,现有的基于案例推理系统的案例检索模型主要有以下几种。



(1) 距离度量法或最近邻算法(KNN)

相似性是衡量对象之间相似度的指标，一般通过计算两个对象在特征空间的距离来获得。在现有 CBR 系统中，大多数案例检索都使用最近邻算法。距离度量以及相似性函数对不相关的噪声特征很敏感，噪声特征导致了 KNN 算法较高的出错率。改进的 KNN 算法使用带权重的距离度量算法，即给数据集的每个属性赋予一定的权重，以权重大小表示属性在特征空间的重要性程度。检索过程与上述类似，称为加权检索。该算法的核心思想是计算案例间的相似度，找出一个或多个最大相似度的案例为其检索结果。

(2) 分类网模型

按抽象与具体层次结构组织案例，检索时采用自上而下的搜索策略，越往下搜索，案例的描述越具体，相似程度也越高。

(3) 归纳索引法

该方法不断从案例的各种组织结构中抽取出最能将该案例与其他案例区别开的成分，并根据这些成分将案例组织成一个类似于判别网络的层次结构。检索时采用判别树搜索策略，基于决策树来实现。归纳法在检索目标有明确定义、且每种目标类型均有足够多的例子进行归纳的情况下，优于最近邻方法，可以根据案例自动确定最佳的用于检索的特征。归纳法的缺点在于需要用大量案例进行归纳，而且归纳学习将增加建立知识库的时间。

(4) 基于知识的深层检索

该方法根据目前已知的知识来决定案例中哪些特征在进行案例检索时是重要的，并根据这些特征来组织和检索。通常将其与最近邻方法结合使用，并利用表层与深层知识的结合，其深层知识包括领域的因果模型，能针对推理过程、方法策略修改的原则对求解结果进行解释。其深层知识避免了不相关案例的检索，甚至可在检索过程中对索引方式进行动态修改，这就使得案例的组织与检索具有一定的动态性。

(5) 神经网络检索方式

神经网络检索方式利用神经网络的诸多优点，建立一个案例库的并行分布式神经网络的表达模型，然后根据用户输入来搜索案例库网络，具有快速、自适应、抗噪声等优点。

(6) 粗糙集检索方式

粗糙集检索方式不需要预先给定某些属性的数量描述，而是直接从给定问题的描述集合出发，通过不可分辨关系和不可分辨类确定给定问题的近似域，从而找出该问题的内在规律。如将粗糙集理论引入 CBR 中，充分利用案例库中冗余属性的简化，形成案例的索引，从而可以根据不同问题按不同索引进行检索并得出结论。

进行案例检索的根本目的就是从众多案例中查找和搜索出与当前问题最相似的案例，实现这一目的的方法就是通过一定的算法进行案例与问题之间的相似性比较，因此，相似性计算方法是进行案例检索的基础，对案例检索过程、检索结果甚至对问题的求解都具有重大的影响。事实上，案例检索过程的本质就是根据一定的相似性度量方法从案例集合(案例库)中查找和搜索出与当前问题最相似案例的过程，因此，该过程表现为一种通过适当的途径进行相似性比较的过程。目前，常用的相似性度量模型主要有下述几种。

(1) 最近邻近相似性度量模型

$$S_{nk} = \frac{\sum_{i=1}^{m} w_i V_{nk}^i}{\sum_{i=1}^{m} w_i}$$

式中，S_{nk} 表示案例 n 和案例 k 的相似性系数；w_i 表示第 i 个属性在整个案例中的权重；V_{nk}^i 表示案例 n 和案例 k 的第 i 个属性相似性系数；m 表示案例的属性个数。

可以看出，最近邻近相似性度量模型有一个假设前提，即认为两个案例之间的属性集是一样的，且两个案例的属性权值完全相同。这样，在应用时案例的表示和权值的赋予均比较简单。

（2）TC 相似性度量模型

$$TC_{nk} = \frac{A^n \cap A^k}{(A^n \cup A^k) - (A^n \cap A^k)}$$

其中，A^n 表示案例 n 的属性全集；A^k 表示案例 k 的属性全集；TC_{nk} 表示案例 n 和 k 之间的相似性系数。

TC 相似性度量模型是一种通过矩阵运算表示的相似性度量关系，它适用于属性可以用二进制表示的应用领域。

（3）改进的 TC 相似性度量模型

$$S_{nk} = \frac{\sum_{i=1}^{m} w(n,i)w(k,i)V_{nk}^i}{\sum_{i=1}^{m}(w(n,i))^2 \sum_{i=1}^{m}(w(k,i))^2}$$

以上是改进的 Tversky 匹配法的相似性度量模型。其中，$w(n,i)$，$w(k,i)$ 分别表示第 i 个属性在案例 n，k 中的权值；V_{nk}^i 表示案例 n 和 k 的第 i 个属性的相似度；m 表示案例 n 和 k 的所有属性的个数；S_{nk} 为其相似度。

改进的 Tversky 匹配法考虑了属性集中的各属性对于两个案例具有不同的权值，对于属性集相同而权值不同的案例可用此法求解相似度。对于案例 k 有的而案例 n 没有的第 i 个属性，令 $w(n,i)$ 为 0，反之亦然。这样就可以求得案例 n 和案例 k 的相似度。

6.3.2 基于规则与案例结合的检索方法

人们解决问题的方法不外乎两种：知识和经验。如果将案例（CBR）视为经验，那么规则（RBR）即为原理性知识。用知识来解决问题是演绎推理的方法，用经验来解决问题是归纳推理的方法。现实中求解问题时，这两种知识显然互为补充。虽然与 RBR 相比，CBR 更符合领域专家的思维过程，但它并不独立于传统专家系统单独存在，而是与规则相辅相成，因为在实际中人们解决问题时，既不是单纯靠经验，也不是单纯靠理论知识，而经常是两者的结合。因为人们的领域知识是有限的或不可靠的，即使是专家也会随着科学研究的深入和发展暴露出一些局限性或错误，所以需要经验来验证结果的正确性；而完全脱离现有的领域知识，只靠经验来解决问题，则是盲目的，所以应有领域知识作指导。理论和经验紧密联系，相辅相成。同一理论可以表现为多种经验，经验当中也必然蕴涵着一定的理论知识。将已有的领域知识同过去的经验结合起来，无疑是一项正确的选择。[5]

一般来说，CBR 系统适用于经验丰富，但知识贫乏的领域；而 RBR 系统适合知识丰富且被系统化描述的领域。因此，将 CBR 和 RBR 合二为一的集成系统将拥有两者的长处，克服两者的不足，从而使整个系统达到更高的智能水平，具有十分重要的意义。两种技术的结合，不仅会大大增强系统的灵活性及其综合推理能力，还会明显减轻案例检索与案例库的负担。

我们将 RBR 融入 CBR，即在 CBR 中增加一个知识库，这个知识库可以包括领域知识，用



来指导相似案例比较或案例修改。检索与当前问题状况类似的案例时,不仅要找到可以帮助解决当前问题的案例,而且这个案例还应该便于修改,以得到当前问题的答案。检索到的相似案例还要依次与当前问题的最终解答相比较以抽取修改规则。搜索案例库中的全部案例是不可能的,因为那样做的话算法复杂度太高。另外,随着时间的推移,案例库中的案例会不断增加。如果把每个系统解决的问题都作为一个新案例加入案例库,那么案例库中必然存在许多"重复"案例,对此可以采用两个办法:一是将"重复"案例完全合成为一个案例保存;二是在重复案例中提出一个这组案例的代表,检索时只检索该代表案例,不搜索其余的。事实上,案例表示、索引、检索、修改规则的提取和案例的存储此时相互联系,知识将它们联系到一起,使其结合成为严整的系统。我们采取这样的案例库结构:所有案例分为两层,即底层是所有的案例;上层是典型案例。每个典型案例都代表了一个或若干个与它极为相似的案例,典型案例间的修改规则至关重要,应该有领域专家的指导。要得到良好的修改规则,对于相似案例间语义差异的衡量至关重要,语义可以由索引词表定义的关键字来描述,索引词表的关键字同样可以用于案例功能及语义网节点的描述。应用修改规则还需要案例评价,即对子系统给出的解答的评价,因为一次修改得到的结果可能不成立,还要进一步修改。有时问题无法与某一个规则的前件匹配,此时需要将问题分别映象到几个规则的前件,最后将几条规则的演绎结果综合成为整个问题的解答。基于层次聚类建立智能化层次案例库,有效地缩小了 KNN 检索时的案例搜索空间,提高了检索效率。

1. 案例索引

案例库的索引(indexing)目标是提供一种案例库的搜索机制,使得在将来的检索中能够快速找出符合需要的案例或案例集。好的索引有助于快速、准确地找到所需案例。索引可以是案例的表面特性或导出特性,表面特性可以从案例描述中直接得出,而导出特性则必须通过一定的推理才能得到。不管是何种特性,最重要的是索引应该是与推理或解决问题的过程自然相关的那些关键字的集合,这些关键字可以将这个案例同其他案例区分开来。

对此,传统的信息检索和数据库系统的索引技术有很好的借鉴作用,但是建立案例索引的原则与传统信息检索和数据库系统的索引概念不同。在传统的索引概念中,一个重要的指导思想是保持存储结构平衡。在 CBR 中,索引用来在某种需要下指出一个独立的案例,关心如何把案例库划分成概念上有用的片段,而不再关心组织结构的平衡。另一点不同之处在于,大多数据库操作是进行精确匹配,而在 CBR 中则是进行相似性匹配。

这里借鉴聚类的概念,采用 C 均值聚类算法,以一个聚类中心代表一种疾病,按疾病的种类个数分割为几个聚类中心,按聚合中心进行索引。

C 均值聚类模型为

设目标函数
$$J=\min\sum_{r=1}^{P}\sum_{i=1}^{m_r}\|X_i^{(r)}-C_r\|^2$$

其中,聚类中心 $C_r=\dfrac{1}{m_r}\sum_{i=1}^{m_r}X_i^{(r)}\left(i=1,2,\cdots,m_r;r=1,2,\cdots,P,\ \sum_{r=1}^{P}m_r=N\right)$;$m_r$ 为属于 r 类的样本(记录)个数;$X_i^{(r)}$ 表示样本 X_i 属于第 r 类;N 为样本(记录)数;P 为聚类中心数($2\leqslant P\leqslant N-1$)。

以下用遗传算法求解动态聚类问题。

步骤 1:染色体的构造。

用 $X=(X_1,X_2,\cdots,X_N)$ 表示染色体结构，$\overline{X}$ 为 $1\times N$ 维行向量，X_k 为第 k 位的基因，则染色体 $X_k\in\{1,2,\cdots,N\}$，使用自然数编码，通过染色体来动态确定聚类数目。

设置遗传算法的相关参数：max_gen 为最大迭代次数，N 为群体大小，length 为染色体长度，P_c 为交叉概率，P_m 为变异概率，P 为初始聚类中心，α 为在适应值函数中的参数。

步骤 2：群体初始化。

```
for i=1 to N do
    for j=1 to length do
        染色体 Xi 的第 p 位基因=random(0,P);
    endfor
endfor
```

步骤 3：计算适应值函数。

```
for i= 1 to N do
    计算染色体 Xi 的目标函数 Ji;
endfor
根据染色体 Xi 的目标函数值 Ji 进行排序;
for i= 1 to N do
    计算个体适应度
endfor
for j= 1 to N do
    计算累积概论
endfor
```

步骤 4：执行选择操作。

步骤 5：执行交叉操作。

步骤 6：执行变异操作。

步骤 7：精英保留策略。

```
if (Minimumof NewGeneration< Minimumof OldGeneration) then
    用父代中最好的个体替换子代中最差的个体;
endfor
```

步骤 8：动态改变聚类数目。

```
If (群体中的最好个体在 M 代内没有改变) then
    P= Popt+ n,并转步骤 2;
else
继续运行;
endif
```

步骤 9：终止条件检验。

```
if (gen< max_gen) then
    令 gen= gen+ 1,并转步骤 3;
  else
      停止计算,输出结果
endif
```



2. 基于规则和案例的检索算法

基于规则和案例的检索算法如下。

步骤1：输入新案例，按照索引生成的有关规则生成新案例的规则描述。

步骤2：判断规则编码RULE_CODE的相似程度，若完全相同，则直接定位案例。若不完全相同，分析属性的相似值是否在阈值范围内；若在阈值内，检索出对应的案例集；若不在阈值内，根据案例的相似度进行匹配，进而检索出相应的案例。若检索出的案例不合理或检索不出案例，转步骤4。

步骤3：检索过程中，保存有关案例学习过程的关键知识和兴趣爱好，存入启发性知识库。

步骤4：选定案例系统涉及的对象，按照改进的最相邻近法计算对象下的相似度衡量。

步骤5：判断是否还涉及其他对象，若是，则取案例系统涉及的下一对象；否则转步骤2。

步骤6：计算出案例系统总的相似度。

步骤7：相似度最高的案例就是结果。

步骤8：判断该案例用户或专家是否满意，若满意，则将该类案例列为候选案例集，结束推理；若不满意，则转到步骤4。

6.4 基于遗传算法的权重优化模型

基于案例的学习是指对于给定的问题描述，系统在案例库中进行检索，寻找最佳匹配的案例。若找到完全匹配的案例，问题就得到解决。若找不到完全匹配的案例，系统则寻找类似案例，再根据此案例和问题描述的差别，对此案例的解进行修正，使之适应当前的问题描述，并将该解加入案例库中以备应用。这既是对系统知识不断积累、补充、修正和更新的过程，也是案例推理和问题求解过程的重要组成部分。

在案例学习过程中，学习的对象通常是系统中出现的某一个具体事件或问题。因此，对具体事件或问题所具有的特征属性进行准确的描述和表达，是案例学习的基本内容，而新案例的建立或对已有案例的修正则是案例学习的主要途径。

在通常情况下，一个案例由特征本身以及该特征所具有的权值两个部分组成。案例的特征属性是决定案例基本性质的主要因素，是考察和判断案例与案例之间、案例与所要解决的问题之间相互关系及其性质的基础，是案例推理的基础和主要依据。而另一方面由于各个特征在案例中所具有的地位不同，对案例性质的影响程度和所起作用也不相同，所以，这些特征需要通过适当的权值配置，才能在案例中发挥与其性质相适应的作用，因此案例特征权值的分配和确定则是建立案例的另一项重要基础工作。

给特征项赋予权重的常用算法是Salzberg于1991年在EACH中使用的方法，该方法的基本思想为：一是增加那些与同类或接近的案例的相似性；二是降低不同类的相邻案例的相似性。当某个分类是正确的情况下，如果特征 f 匹配，那么它的权重 W_f 就可以增加 Δf，如果特征 f 不匹配，那么它的权重 W_f 将被降低 Δf。当某分类不正确时，不匹配特征项的权重将增加，而匹配特征项的权重将减少。[6]

案例学习与特征属性有着直接关联，我们将结合鱼病诊断推理系统的具体特点，提出一种基于GA的案例权值修正算法，以通过对特征权值的调整和优化来实现案例学习的目的。

6.4.1 适应度函数数学模型的建立

遗传算法在进化过程中基本上不利用外部信息,仅以适应度函数为依据,利用种群中每个个体的适应度值来进行搜索。因此适应度函数的选取至关重要,直接影响到遗传算法的收敛速度以及能否找到最优解。对案例特征权值向量优劣的评定指标通常采用检索精度,也就是遗传算法中的适应度函数的度量,因此遗传算法的优化目标就是寻找最优化的特征向量从而使适应度函数最大。[7]

将案例分为测试案例集 TEST 和学习案例集 REF。测试案例集为测试 CBR 系统检索精度的案例集合,学习案例集为遗传算法学习的案例集合,也是 CBR 系统运行时案例库。这里采用 KNN 的模型,它被用作检索精度,用在遗传中每个染色体的适应度函数上。对于学习案例集 REF 和测试案例集 TEST,令 $\mathrm{ref}[i]\in \mathrm{REF}$, $\mathrm{test}[j]\in \mathrm{TEST}$, $i=0,1,\cdots,N_{\mathrm{R}}$, $j=0,1,\cdots,N_{\mathrm{T}}$,其中 $\mathrm{ref}[i]$, $\mathrm{test}[j]$分别表示学习集和测试集中的第 i,j 个案例,N_{R} 为学习集中的案例数,N_{T} 为测试集中的案例数。

一个染色体代表一个权矢量,一个染色体由多个基因组成,每个基因表示的就是单个基因项的权重。一个染色体所含的基因个数,等同于案例集中独立特征项的个数。对于每个权矢量$\boldsymbol{\omega}[i]$,找到与每个测试案例 $\mathrm{test}[j]$距离最近的学习案例,利用所有的测试案例与它们在学习集中最近邻的距离 distan ce(j,k)之和作为适应度函数 $\mathrm{fit}[i]$,并以此评估每一个权矢量的遗传能力。其模型如下:

$$\max\ \mathrm{fit}[i] = 1-\sum_{j=0}^{N_{\mathrm{R}}}\sum_{k=0}^{N_{\mathrm{T}}} d(\mathrm{test}[j],\mathrm{ref}[k])$$

$$\mathrm{s.t.}\ d(\mathrm{test}[j],\mathrm{ref}[k]) = \left[\sum_{f}\boldsymbol{\omega}[i]_f\times d(\mathrm{test}[j]_f-\mathrm{ref}[k]_f)^2\right]^{1/2}$$

其中,$d(\mathrm{test}[j],\mathrm{ref}[k])$实际上就是表示第 j 个测试案例与其最近的参考案例的距离,f 表示数据库中案例的第 f 个相关特征,$\boldsymbol{\omega}[i]$表示第 i 个权矢量的值,对应的是第 f 个特征项的权值。

6.4.2 自适应的遗传算子

遗传算法本质上是一种群体寻优的迭代过程,它从一个随机初始解群出发,通过选择、交叉、变异等操作,产生性能更优的下一代解群,这一过程一直持续下去,直到产生满足约束条件的最优个体为止。[8]

遗传算法的参数中交叉概率 P_{c} 和变异概率 P_{m} 的选择是影响遗传算法行为和性能的关键所在,直接影响算法的收敛性,在标准遗传算法的运行过程中,交叉和变异概率是固定的,P_{c} 越大,新个体产生的速度就越快,而 P_{c} 过大时遗传模式被破坏的可能性也越大,使得具有高适应度的个体结构很快就会被破坏;但是如果 P_{c} 过小,会使搜索过程缓慢,以至停滞不前。对于变异概率 P_{m},如果 P_{m} 过小,就不易产生新的个体结构、因而不容易跳出局部极值点;如果 P_{m} 取值过大,那么遗传算法就变成了纯粹的随机搜索算法。针对不同的优化问题,需要反复实验来确定 P_{c} 和 P_{m},这是一件繁琐的工作,而且很难找到适应于每个问题的最佳值。Srinivas 等提出一种自适应遗传算法(AGA),P_{c} 和 P_{m} 能够随适应度自动改变[9]。当种群个体适应度趋于一致或者趋于局部最优解,使 P_{c} 和 P_{m} 增加;而当群体适应度比较分散时,使 P_{c} 和 P_{m} 减少。同时,对于适应值高于群体平均适应值的个体,对应于较低的 P_{c} 和 P_{m},使该解



得以保护进入下一代；而低于平均适应值的个体，相对应于较高的 P_c 和 P_m 使该解被淘汰掉。因此，自适应的 P_c 和 P_m 能够提供相对某个解的最佳 P_c 和 P_m。自适应遗传算法在保持群体多样性的同时，还保证了遗传算法的收敛性。

在自适应遗传算法中，P_c 和 P_m 按如下公式进行自适应调整：

$$P_c=\begin{cases}\dfrac{k_1(f_{\max}-f')}{f_{\max}-f_{\text{avg}}}, & f\geqslant f_{\text{avg}}\\ k_2, & f<f_{\text{avg}}\end{cases}$$

$$P_m=\begin{cases}\dfrac{k_3(f_{\max}-f)}{f_{\max}-f_{\text{avg}}}, & f\geqslant f_{\text{avg}}\\ k_4, & f<f_{\text{avg}}\end{cases}$$

其中，$f_{\max}$ 为群体中最大的适应度值；f_{avg} 为每代群体的平均适应度值；f' 为要交叉的两个个体中较大的适应度值；f 为要变异个体的适应度值。

这里，只要设定 k_1,k_2,k_3,k_4 的取值范围为(0,1)区间，就可以进行自适应调整了。

当适应度值低于平均适应度值时，说明该个体是性能不好的个体，于是对它采用较大的交叉率和变异率；如果适应度值高于平均适应度值，说明该个体性能优良，对它就根据其适应度值取相应的交叉率和变异率。可以看出，当适应度值越接近最大适应度值时，交叉率和变异率就越小；当等于最大适应度值时，交叉率和变异率的值为零。这种调整方法在群体处于进化后期时比较合适，但对于进化初期不利。因为进化初期群体中的较优个体几乎处于一种不发生变化的状态，而此时的优良个体不一定是优化的全局最优解，这容易使进化走向局部最优解的可能性增加。可以进一步改进，使群体中最大适应度值的个体的交叉率和变异率不为零，分别提高到 P_{c2} 和 P_{m2}，这就相应地提高了群体中表现优良的个体的交叉率和变异率，使得它们不会处于一种近似停滞不前的状态。

经过上述改进，P_c 和 P_m 计算表达式如下：

$$P_c=\begin{cases}P_{c1}-\dfrac{(P_{c1}-P_{c2})(f'-f_{\max})}{f_{\max}-f_{\text{avg}}}, & f'\geqslant f_{\text{avg}}\\ P_{c1}, & f'<f_{\text{avg}}\end{cases}$$

$$P_m=\begin{cases}P_{m1}-\dfrac{(P_{m1}-P_{m2})(f_{\max}-f)}{f_{\max}-f_{\text{avg}}}, & f\geqslant f_{\text{avg}}\\ P_{m1}, & f<f_{\text{avg}}\end{cases}$$

通常情况下，上式中 $P_{c1}=0.9$，$P_{c2}=0.6$，$P_{m1}=0.1$，$P_{m2}=0.001$。

对于选择算子，我们采取了排序选择的方法，为了保证每一代的优良个体不被破坏，采用精英选择策略，使它们直接复制到下一代中。

6.4.3　基于遗传算法 GA 的权重优化算法

基于遗传算法 GA(KNN)的权重优化算法的主要步骤如下。

步骤 1：随机产生 m 个案例特征权值向量初始群体 $W=(w_1,w_2,\cdots,w_m)$。

步骤 2：将测试案例集 T 中的案例分别输入 CBR 系统进行检索，根据适应度函数 fitter[i]($i=1,2,\cdots,n$)计算出和测试案例相似度。

步骤 3：根据权矢量的适应值，按从小到大的顺序排序。

步骤 4：如果权矢量适应值小于父代的权矢量适应值，那么用子代替代父代个体；并保留最优个体。

步骤 5：对染色体按 P_c 进行个体交叉操作。

步骤 6：对染色体按 P_m 进行个体变异操作。

步骤 7：判断 GA 的收敛条件(如最大进化代数)是否满足。如满足，则停止，输出当前代中适应度最小的特征权值向量作为优化特征权值向量。否则转步骤 8。

步骤 8：对优胜的特征权值向量进行复制、交叉、变异等操作，产生下一代特征权值向量群体，重复步骤 3 到步骤 8。

需要注意的是，随着新的案例不断增加，如果不采取适当的措施，将会使案例库变得十分庞大，从而其推理效率势必受到影响。因此当案例被加入案例库时，通常遵循以下步骤：

(1) 若案例库中已有相同案例，则弃之。

(2) 替换案例库中的旧案例。

(3) 与案例库中的旧案例合并，形成一个新案例。

(4) 直接作为新案例独立地加入案例库中。

6.5 案例检索模型在鱼病诊断系统中的应用

6.5.1 鱼病诊断推理系统案例集合的结构特点

本节针对鱼病诊断系统案例集合的结构特点，将 CBR 与 RBR 相结合的案例检索方法应用于该系统，以此探讨该方法在鱼病诊断系统中的应用。在我们研究的鱼病诊断系统当中，系统本身以及研究对象存在着一些特殊性和复杂性使得系统中的案例和问题都表现出明显的特点，主要表现在以下方面：

(1) 案例集合的非同构性。鱼病诊断系统的任务是建立一个鱼病诊断与防治平台，由于不同的鱼病在症状、疾病原因、疾病及对疾病的描述和表示上都会有所差别，因此，针对不同的诊断对象的各种疾病现象所建立的案例，在特征属性上也会反映出这种差别，即在不同案例中使用的特征在种类和数量上不同；同一特征对不同案例的影响程度和所起的作用大小也可能不同，因此，同一特征在不同案例中的权重也将发生变化，使得特征的权重表现出局部化特征。所以，在一般情况下，该系统中的案例集合表现出很强的非同构性特征，比如草鱼烂鳃病的症状特征主要表现在头部、鳃部，而草鱼出血病的症状特征主要表现在头部、肌肉、鱼鳍、鳃部，因而各种疾病的特征权重因疾病种类的不同而表现出很强的非同构特征。

(2) 由于鱼病诊断系统对几种特征同时进行检测和诊断，并且系统案例集合具有非同构性特征，故由一次所获数据得出的特征值以及这些特征值之间不同方式的组合，都可能表示系统中某一事件或疾病的发生与出现，都有可能在案例集合中存在与之匹配的案例。因此，对同一组特征值，在检索过程中，有可能查找到多个不同案例与其相匹配。

6.5.2 基于规则与基于案例推理结合的应用

由于鱼的品种很多，鱼病症状又比较复杂，故很难直接给出正确全面的分类，从而对检索效率提出了挑战。根据鱼病诊断的实际特点，以及这几年中国农业大学知识管理中心对鱼病专家系统的规则及案例所积累的知识和经验，我们提出并构造了按层次智能分类的抽象规则



集和具体案例集,并将该方法应用于鱼病诊断推理系统。

该方法的思路是,将原始案例库分割为聚合中心库(即抽象规则库)和案例库,首先由抽象规则库组成第一级索引,借鉴聚类的概念,采用 C 均值聚类算法,以一个聚合中心代表一种疾病,在本系统中以 9 种常见疾病为例,规则库由各个聚合中心构成,聚合中心是某一类鱼病的代表案例。这样在鱼病诊断系统中,可通过 RBR 子系统建立当前问题的索引,主要是利用已经积累起来的领域知识,对例子空间直接进行划分,形成典型案例库,这样可以最大限度地减少典型案例库中的案例数目,从而降低存储需求,提高检索速度并使得抽取的典型案例更具有代表性,缩小了案例库的搜索范围;其次由具体的案例库组成第二级索引,它包括属于同一类的各个案例,一个类库对应于一个聚合中心。

系统运行时,根据分类结果建立了案例库,采用两级索引。当向系统输入新的鱼病诊断症状时,首先在规则集中进行归类,由此得出和目标案例相似的一类源案例,初步确定疾病,实质上这是正向推理。这时根据初步确定的疾病,在案例库中根据分类标识进行匹配,若完全匹配,则可确诊疾病,并给出相应的治疗方案。若不完全匹配,则根据分类标识用 KNN 算法找出和目标疾病相似度最大的案例,并进行适当修正,重新进行检索,该检索是在大幅度缩小的搜索空间里进行的,检索效率明显提高,这一步实际上是进行反向推理,在案例库中证实上一步得出的结论是否正确。

该方法的最大特点是检索速度与库的大小呈非线性关系,即库的大小不直接与检索速度相关,克服了传统案例推理中的检索速度与库的大小存在着线性关系这个缺陷。该方法中学习和匹配是两个独立的过程,即进行鱼病的案例匹配时不进行鱼病的案例学习,而进行案例学习时不进行案例的匹配。系统的学习是需要时间的,学习一次后系统可以使用很长的一段时间,但每一次匹配所需的时间很短,远小于传统的基于案例的检索模型的一次匹配所花费的时间,对于大型库且实时性要求高的应用领域也能够满足需求。而且,在实际的应用中,学习和匹配可以并行工作,克服了学习时间长的缺点。

本系统使用归纳方法基于案例历史记录建立决策树模型,然后用该模型来解决问题。归纳学习要求结构化的数据组织,一个标准的关系数据库易于映射到这种数据组织上,这样使用者可以自己定义描述案例的字典。给定数据模型、目标类和案例数据,系统可自动产生决策树,从而可以从案例历史记录中提取出有关的决策知识。

鱼病诊断推理系统的一个目标就是实现鱼病的早期预防和诊断,为用户减少鱼病诊断的所有费用。我们开发的鱼病诊断系统,结合使用了归纳和基于案例的推理技术。这个系统帮助用户执行在线疾病诊断,可更快和更好地给出建议。整个系统用于:

(1)减少了解决那些对专门技术人员要求很高的问题所花费的鱼病诊断时间,避免了鱼病诊断与防治的延迟。

(2)将诊断费用降至最低。

(3)减少了鱼病诊断推理系统错误诊断的次数。这样的错误会导致向用户提供错误的防治方法,并且由于用户不得不等待正确的诊断结果,给鱼病诊断带来了延误。由于该系统用于为更多远程网络用户服务,这个问题变得更加重要。

(4)选择最具辨别力的测试来识别疾病,将诊断过程形式化。

(5)记录技能熟练的鱼病专家的经验并整理成文档,以建立相应的记忆并帮助把专家的

诀窍传递给新手,这样可节省新手的学习费用。在新手无法获得关于这些疾病的经验的情况下,这一点特别重要。

鱼病诊断中典型案例可协助解决60%至70%的问题。因此鱼病诊断推理系统包含了一个典型案例库,它可在网络上共享。一个内部组织(由鱼病专家组成)在案例加入典型案例库之前会进行审查,以控制库中案例的质量。典型案例库随着案例的增加不断丰富。

为了适应用户的思维方式,这个系统整合了CBR和归纳学习技术。当有新案例输入时,系统用一棵事先在数据库里建立的归纳树进行预处理,以解决最常见的问题。

在会诊过程中,归纳学习产生了一棵新的疾病树,它与标准疾病树可能有所不同,该疾病树是根据实际观测到的疾病生成的,并且可以根据新出现的症状进行更新。检索这样一棵疾病树时,操作者要回答与原始疾病症状相关的一些问题。在会诊过程的最后,系统按照相对频率为操作者提供一份可能的解决方案列表,然后检索出支持结论的案例,由操作者通过浏览可确定解决方案。

有75%的诊断问题都可以在这一步得到解决。有意将归纳树的尺寸设置得比较小,以使诊断推理所用的时间比较短。如果问题还没有解决,例如还有多个诊断悬而未决,系统会自动启用"动态归纳"产生一个相关问题的列表,还会自动产生一份包含目前诊断结论和问题的报告,并向用户显示。用户在看到确切的诊断结论之前要进一步回答一些问题,然后系统检索出以前的问题,并根据用户对问题的回答将缺少的信息补充完整。接着系统会在相关案例中进行最近邻搜索,并把最接近的案例及其相关诊断返回给用户,同时给出防治的方法和策略。

6.6 小结

本章根据鱼病诊断的特征及内容,对基于案例的推理方法进行了系统研究,并将其应用于鱼病诊断系统,可有效地解决鱼病诊断问题,并提高系统的运行效率。

案例学习是CBR中的关键技术问题之一,它的性能和效率对保证系统对问题求解的有效性、准确性有着至关重要的影响,而案例的特征项权重对检索的质量与速度都起到了重要的作用。本章提出了一种利用遗传算法来完成特征项权空间直接搜索的算法,由遗传算法进行特征权值的优化。实践证明:该方法可防止检索出的相似度系数最大的案例并非最佳案例,即KNN收敛不到最佳解,提高了检索质量。

参考文献

1. Noh JB. A case-based reasoning approach to cognitive map-driven tacit knowledge management[J]. Expert Systems with Application. 2000(19):249～259
2. Deng PS. Using case-based reasoning approach to the support of ill-structured decisions. European Journal of Operational Research[J]. 1996 (93):511～521
3. Garrell JM,Guiu I. Automatic diagnosis with genetic algorithms and case-based reasoning[J]. Artificial Intelligence in Engineering. 13(1999):367～372
4. Liao TW. A case-based reasoning system for identifying failure mechanisms[J]. Engineering Applications of Artificial Intelligence. 2000 (13):199～213



5. 张晓莉，杨杰，吕永. 诊断推理中人工神经网络与基于案例推理的结合[J]. 上海铁道大学学报，2000(6)：6～11
6. Salzberg S. A nearest hyperrectangle dearning methord[J]. Machine Learning，1991(6)：251～276
7. 黄炎，蒋培等. 基于可调变异算子求解遗传算法的欺骗问题. 软件学报，1999，10(2)：216～219
8. 邢文训，谢金星，现代优化计算方法[M]. 北京：清华大学出版社，2005
9. Srinivas M，Patnaik LM. Adaptive probabilities of crossover and mutation in genetic algorithm[J]. IEEE Trans on System，Man and Cybernetics，1994，24(4)：656～667

第 7 章

基于 CBR 和 RBR 协同推理的鱼病诊断推理模型

人们在解决实际问题时,既不单纯依靠经验,也不单纯依靠理论知识,而是经常依靠两者的结合。针对这种情况,本章介绍了基于 CBR 和 RBR 协同推理的鱼病诊断推理模型,包括鱼病诊断案例索引模型和案例检索模型。

7.1 鱼病诊断的 CBR 与 RBR 协同推理机制

鱼病诊断问题十分复杂,现实中大都依靠经验知识进行诊断。以往的鱼病诊断专家系统多采用 RBR(基于规则的推理)方法。这种方法应用广泛也很成熟,适合处理知识丰富的领域。对于鱼病诊断问题来说,领域知识不完备使得 RBR 不能发挥出应有的作用。CBR(基于案例的推理)系统依赖于特殊情况和不断积累的经验,适用于经验丰富而知识贫乏的领域,而且 CBR 的推理方式更类似于人类处理问题的思维方式。但是 CBR(基于案例的推理)面对越来越庞大的案例库时,其推理效率也会降低。由于 RBR 和 CBR 的不同侧重点,因此二者可以互相配合、互相补充。

将 CBR 和 RBR 合二为一的集成系统,不仅会大大增强系统的灵活性及其综合推理能力,还会减轻案例检索与案例库的负担。因此,我们将 CBR 与 RBR 融合起来,以 CBR 为主、RBR 为辅进行协同推理,如图 7-1 所示。

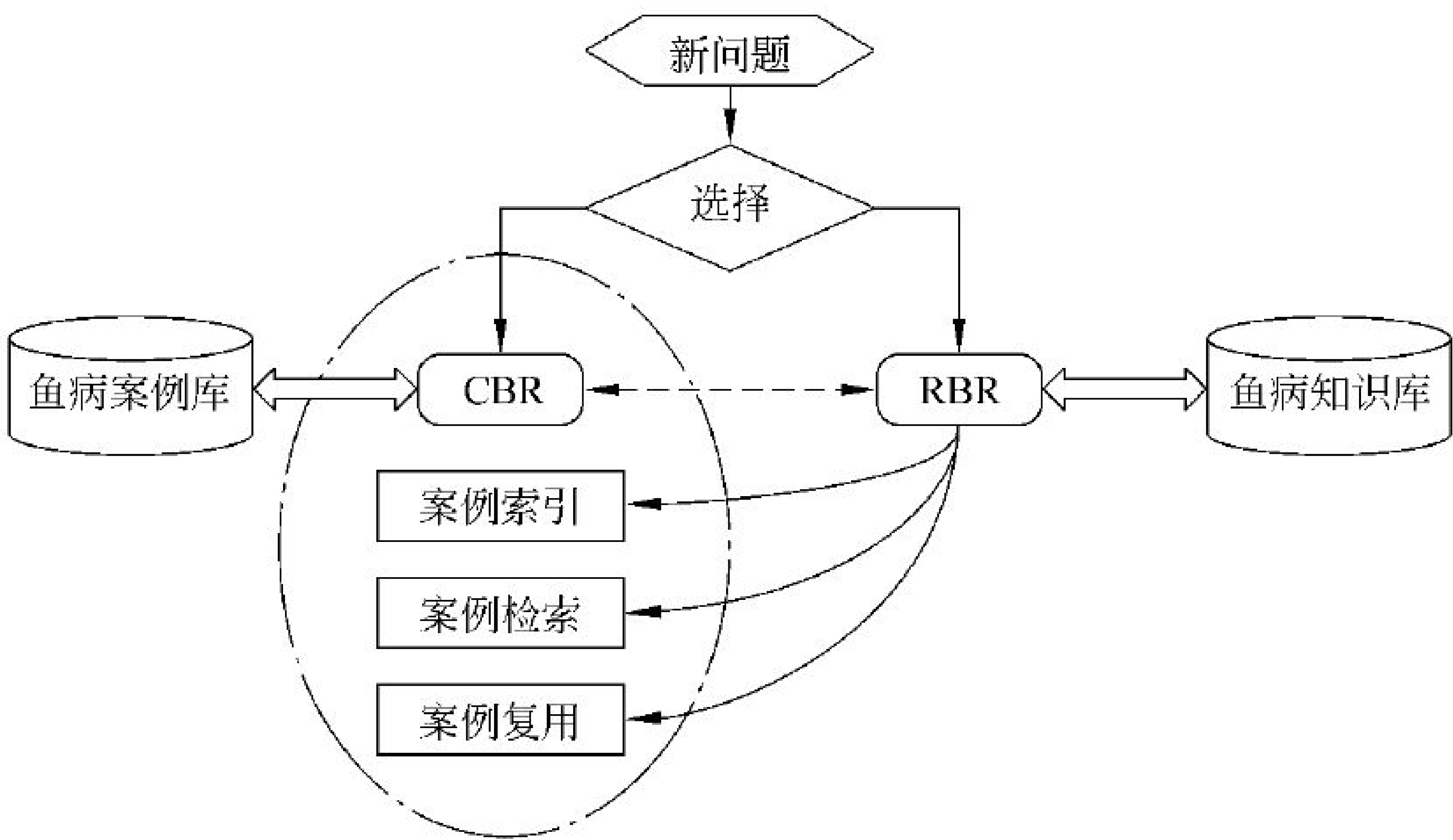


图 7-1　鱼病诊断的 CBR 与 RBR 协同推理机制

我们设计的 CBR 与 RBR 协同推理模型中,包括以下四个方面。

(1) 案例知识与规则知识的互相转换。实际上,每个案例都隐含着规则,同类案例对应着



某一条规则。随着时间的推移,案例库中的案例不断增加,如果把每个系统解决的问题都作为一个新案例添加到案例库中,那么案例库中必然有许多“重复”的案例。对鱼病案例进行聚类分析,将具有相似结构的案例泛化出典型案例,每个典型案例都代表了一个或若干个与其极为相似的案例,并可转换成规则,然后按照规则知识对案例进行求精,一方面要保证案例库的覆盖面不减少,另一方面要减少案例库中的冗余和不太可靠的知识。

(2) RBR 对 CBR 的全面支持。在 CBR 中的几个重要环节中,包括案例索引、案例检索和案例复用,都运用了 RBR 的方法和原理。在案例索引中,由鱼病典型案例和鱼病规则知识构成第一级案例索引,然后对新案例进行规则判定,使每一个案例定位到目标案例库,再进行相似性度量,这样大大减小了案例检索空间。在案例复用中,以鱼病诊断规则为修改规则,对待定鱼病诊断结果进行修改,以适应新案例的要求。

(3) CBR 与 RBR 分别推理。在用户提交鱼病症状表现后,由用户选择是进行案例推理还是规则推理,也可以分别进行推理,再比较诊断结果作出决策。对于鱼病诊断来说,有很多疾病的症状表现具有唯一性和排他性,针对这些疾病利用规则推理的效率是很高的。而对于一些具有很多相似症状特征的疾病则可通过案例推理来进行诊断。

(4) 以 CBR 为主、RBR 为辅进行推理。我们所设计的推理模型先进行案例推理,若顺利得到诊断结果则将其输出,否则再进行规则推理,直到得到问题的解答。

在实际中,基层鱼病医生在解决问题时,既不单纯依靠经验也不单纯依靠理论知识,而是将两者相结合。因为对鱼病领域知识的认识是有限的或不确定的,需要经验来验证方法的正确;而完全脱离领域知识,只靠经验来解决问题,则显得盲目,因此需要领域知识的指导。理论知识与实际经验是紧密联系、相辅相成的。同一理论可以表现为多种经验,经验当中必然蕴涵着一定的理论知识。将理论知识与实际经验结合起来,无疑将获得更好的问题解决方案,这也是我们将 CBR 与 RBR 结合起来使用的目的。

7.2　鱼病诊断的案例检索模型

在鱼病诊断系统中,案例检索的目的就是根据鱼所表现出的特征描述,按症状特征分配方案在鱼病案例库中检索出的具有最相似特征的诊断案例,获得当前问题的解决方案。该最相似案例可以是一个最佳案例,也可以是一组相似案例。综合各相似案例的解决方案,形成最终解决方案作为对鱼病诊断的参考。其中相关的案例检索模型是鱼病诊断推理系统实现的关键部分。我们所研究的鱼病诊断知识的案例检索包括四个步骤:案例特征辨识、案例初级匹配、案例相似性度量和案例复用,如图 7-2 所示。

系统运行时,养鱼户通过电话提交鱼病症状信息,坐席人员对这些症状信息进行特征辨识,将辨识后的症状信息提交给系统。系统根据两级索引,首先在第一级索引的规则集合中进行归类,由此得出和目标案例相似的一类源案例群,初步确定疾病,实质上这是正向推理;再根据初步确定的疾病,在案例库中根据分类标识进行匹配,若完全匹配则确诊疾病,并给出相应的治疗方案;若不完全匹配则根据分类标识用最近邻算法找出和目标疾病相似度最大的案例,并进行适当修正,重新进行检索。该检索在大幅度缩小的案例空间内进行,检索效率明显提高,这一步实际上是进行的反向推理,在案例库中来证实上一步得出的结论是否正确。



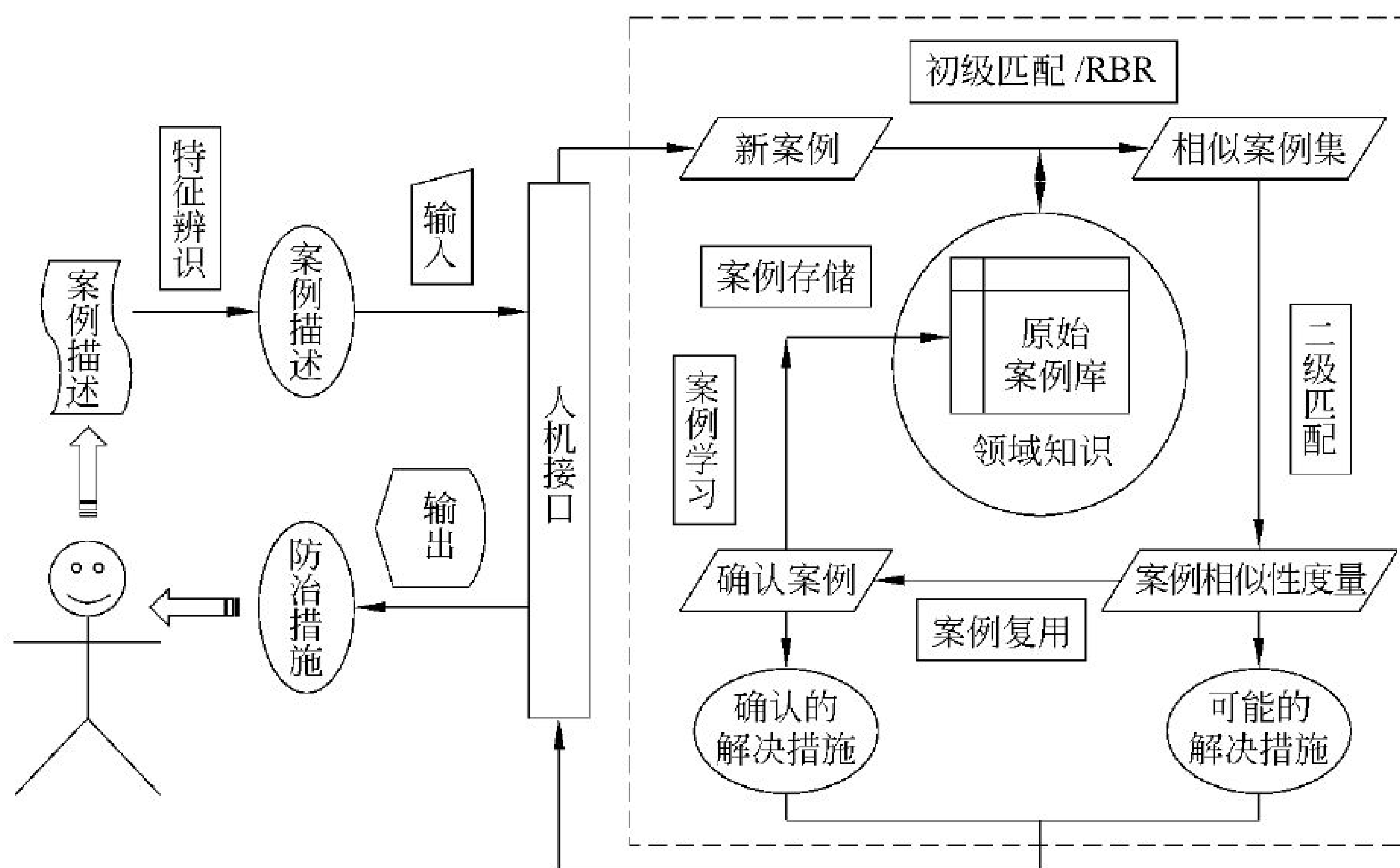


图 7-2　鱼病诊断案例检索模型

7.2.1　鱼病症状特征辨识

本系统面向呼叫中心平台，鱼病案例描述由养鱼户通过电话传递给坐席人员，再由坐席人员通过计算机输入设备将鱼病案例描述输入系统。这一过程中很重要的一个环节就是鱼病案例描述的特征辨识，即由坐席人员对养鱼户给出的鱼病案例描述进行特征辨识和获取，提取出标准的鱼病症状信息。

系统设计过程中，对鱼病诊断案例知识进行了规范化处理和形式化表示，而养鱼户所习惯的是生活语言，这就产生了冲突，需要通过人工或机器智能的方法来进行特征辨识，将养鱼户提交的生活式语言描述与规范化、形式化的鱼病诊断知识描述联系起来。

我们采取人工辅助的半自动方式进行鱼病案例描述的特征辨识，分为两种情况：其一，若可以从用户对鱼病案例描述中直接获得特征属性，则由坐席人员直接将特征属性提交到系统；其二，若不能直接获得，则由坐席人员对用户提交的鱼病案例描述进行分析理解，提取出关键词，将关键词输入到系统的特征辨识辅助模块进行检索，检索出相关特征属性，再从中提取出相符的特征属性，如图 7-3 所示。

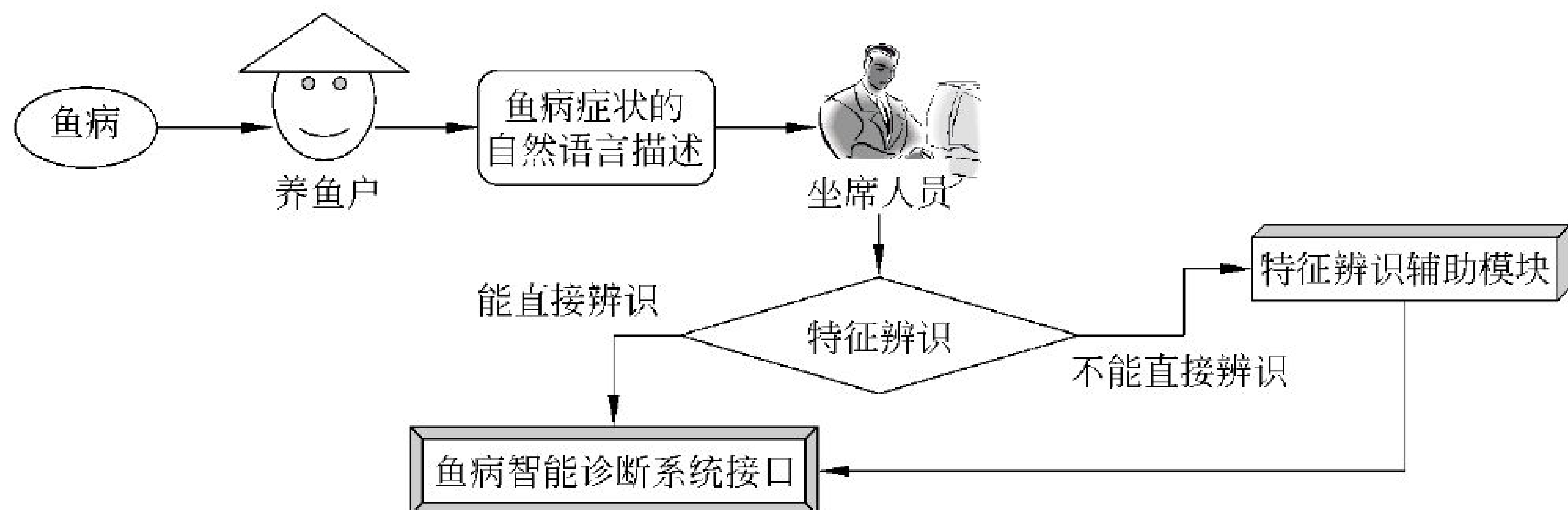


图 7-3　鱼病症状的特征辨识

举例而言，养鱼户提交鱼病症状描述如下：几乎不摄食，逐渐有死亡个体，有的腹部肿胀，肛门红肿，有黄色液体流出，肠道充血，无食物，有脓液，部分有腹水。由此，参照鱼病症状特征



标准值，通过特征辨识提取出鱼病案例症状特征属性：(厌食，死亡，腹部肿胀，肛门红肿，肛门有黄色液体流出，肠道充血，肠道无食，肠道有脓液)。鱼病部分症状特征标准值如表 7-1 所示。

表 7-1　鱼病部分症状特征标准值

部位	症状表现	部位	症状表现	部位	症状表现
现场	离群独游	体表	体表见虫体	鳃部	鳃盖略鼓
现场	游动缓慢	体表	体表有气泡	鳃部	鳃内有白色囊状物
现场	池边打转	体表	鳍上有气泡	鳃部	鳃丝充血
现场	溜边	体表	鳍出血	鳃部	鳃丝暗红
现场	狂游	体表	死鱼胀肚	内脏	腹腔积水
现场	跳跃	体表	眼球突出	内脏	肠道积水
现场	泛坑	体表	鳞片脱落	内脏	肠道充血
现场	浮头	体表	鳞片竖立	内脏	肠道有脓液
现场	痉挛	体表	鳞囊积水	内脏	肠道无食
现场	鱼体失去平衡	体表	腹部肿胀	内脏	肠道有气泡
现场	厌食	体表	腹部充血	内脏	肠道发黄
现场	摄食量大	体表	肛门红肿	内脏	肠道充满食物
现场	死亡	体表	肛门有黄色液体流出	内脏	肝脏暗红
现场	水泛白沫	鳃部	鳃盖有气泡	内脏	肝脏发白
现场	水有异味	鳃部	鳃丝腐烂	内脏	肝脏水肿
体表	鱼体瘦弱	鳃部	鳃丝多黏液	内脏	肝脏发黄
体表	体表发黑	鳃部	鳃盖充血	内脏	肝脏萎缩
体表	体表出血	鳃部	鳃盖张开	内脏	脾脏暗红
体表	体表多黏液	鳃部	鳃部见虫体	内脏	脾脏发黑
体表	体表腐烂	鳃部	鳃丝暗红	内脏	脾脏萎缩
体表	鱼体上有灰白色点状或瘤状胞囊	鳃部	鳃上有白膜		
		鳃部	鳃丝肿胀		

7.2.2　鱼病案例初级匹配

初级匹配是指从案例库中找到一组与当前问题相关的候选案例，就是按照一定的索引策略，先从众多案例中初步筛选出对当前问题可能或潜在有用的案例集，以缩小检索范围，提高查找速度。[1,2]

这里涉及的鱼病案例初级匹配采用模板检索的方法，同时融合 RBR 的方法，即根据系统用户提供的具有特定性质的模式实现检索，并在后台脚本程序设计中加入 RBR 的“IF_THEN”规则推理语句，在案例库中查找所有符合的案例库，并将新案例定位到符合的案例库，其实现类似于关系数据库中的 SQL 查询。所依据的检索模板系通过对鱼病理论知识和案例知识进行分析、整理和融合而成。其中的症状特征表述如表 7-2 所示。

表 7-2　鱼病症状特征

疾病	症状特征描述
竖鳞病	游动缓慢，呼吸困难，身体倒转，腹部向上，有死亡，眼球突出，鱼体呈松球状，皮肤充血，鳍基充血，腹部膨胀，腹腔内积水，鳞的基部水肿，鳞片竖起，用手指在鳞片上稍加压力，渗出液就从鳞片基部喷射出来，鳞片也随之脱落，脱鳞处形成红色溃疡
赤皮病	有死亡，体表出血发炎，鳍基充血，鳍条末端腐烂似一把破扇子，眼球突出，鳞片脱落
气泡病	游动缓慢，游动不正常，身体失去平衡，头朝下，尾朝上，鱼苗死亡，体表有气泡，鳍上有气泡，肠道有气泡
烂鳃病	厌食，鳃丝挂泥，离群独游，鳃丝腐烂，浮头，鳃丝发黑，行动迟缓，鳃丝充血，浮于池边，鳃丝发白，鳃丝肿胀，鳃盖充血，鳃丝黏液多
肠炎病	吃食不好，不吃食，有死亡，腹部肿胀，肛门红肿，肛门外突，肛门有黄色液体流出，肠道充血，肠道积水，肠道有脓液
水质不良	吃食不佳，厌食，浮头，有死亡，鱼体瘦弱，水浑，鳃丝暗红，鳃丝红肿，鳃部黏液增多，鳃边缘有增生物，肠道无食，肠道充血，肝脏暗红，肝脏发黄，腹部膨大，腹腔积水，肠道有气体，肠道有脓液
缺氧	浮头，死亡，泛坑，厌食，水发白，鳃丝贫血，鳃丝暗红，体表发黑，鳍条出血，肠炎
氨氮中毒	扎边，打转，身体痉挛，游动缓慢，浮头，死亡，鳃丝有黏液，体色发黑
寄生虫	不吃食，死亡，游动缓慢，浮头，狂游，跳跃，水体透明度高，鱼体瘦弱，鳃丝充血，鳃丝有分叉，鳃丝肿胀，鳃盖张开，鳃丝红肿，鳃盖略鼓，鳃弓内有白色囊状物，鳃可见透明虫体，鱼体上有灰白色点状或瘤状胞囊，鱼体黏液多，鱼体有虫体，体表寄生着许多鱼鲺，尾柄硬棘发红，体表寄生有孢子虫囊，肠道有寄生虫

症状表现集合 Symptom$=\{s_1,s_2,\cdots,s_m\}$，鱼病集合 Disease$=\{d_1,d_2,\cdots,d_n\}$，以向量分别表示为 $\boldsymbol{S}=(s_1,s_2,\cdots,s_m)$，$\boldsymbol{D}=(d_1,d_2,\cdots,d_n)$，由它们组合成鱼病-症状矩阵为 **S_D**$=|\boldsymbol{S},\boldsymbol{D}|$，检索信息的集合为 $S_X=\{s_1,\ s_2,\cdots,\ s_x\}$。

在 **S_D** 中依次检索集合 S_X 的元素，即 $s_y(1\leqslant y\leqslant x)$，检索得到目标鱼病集合 Disease$_y$。

$$\text{While } y\leqslant x \text{ and } |\text{Disease}_y|\geqslant 1,\quad \text{则返回 Disease}_z=\text{Disease}_y$$

$$\text{While } y\leqslant x \text{ and } |\text{Disease}_y|=0,\quad \text{则返回 Disease}_z=\text{Disease}_{y-1}$$

通过初级匹配，检索出目标 Disease$_z$，进入下一步对每个具体案例进行相似性度量。例如前面表中所辨识出来的特征属性，通过初级匹配可知，该案例描述属于肠炎病一类，即定位到肠炎病案例库。

在上述的匹配过程中，考虑到某些鱼病具有特有症状，即症状 B 只有鱼病 A 才会显现，其他的鱼病不会显现，那么，症状 B 就是鱼病 A 所特有的症状。即当用户提交的症状描述中出现症状 B，至少可以确定病鱼患有鱼病 A，至于是否患有其他鱼病还要按照其他症状进行推理。

由于鱼病症状描述具有很大的模糊性，在匹配过程中造成困难。针对这个问题，采取按照提交的症状描述逐一“扫描”的方式，即按照每一条提交的症状描述逐一检索规则库，每次检索都会产生一组疾病集合，下一次检索在前一次产生的疾病集合中再次检索，直到检索到最少数目的疾病集合。根据实践验证，一般在“扫描”5 次以内就可以检索到目标疾病。鱼病案例的初级匹配的示意图如图 7-4 所示。



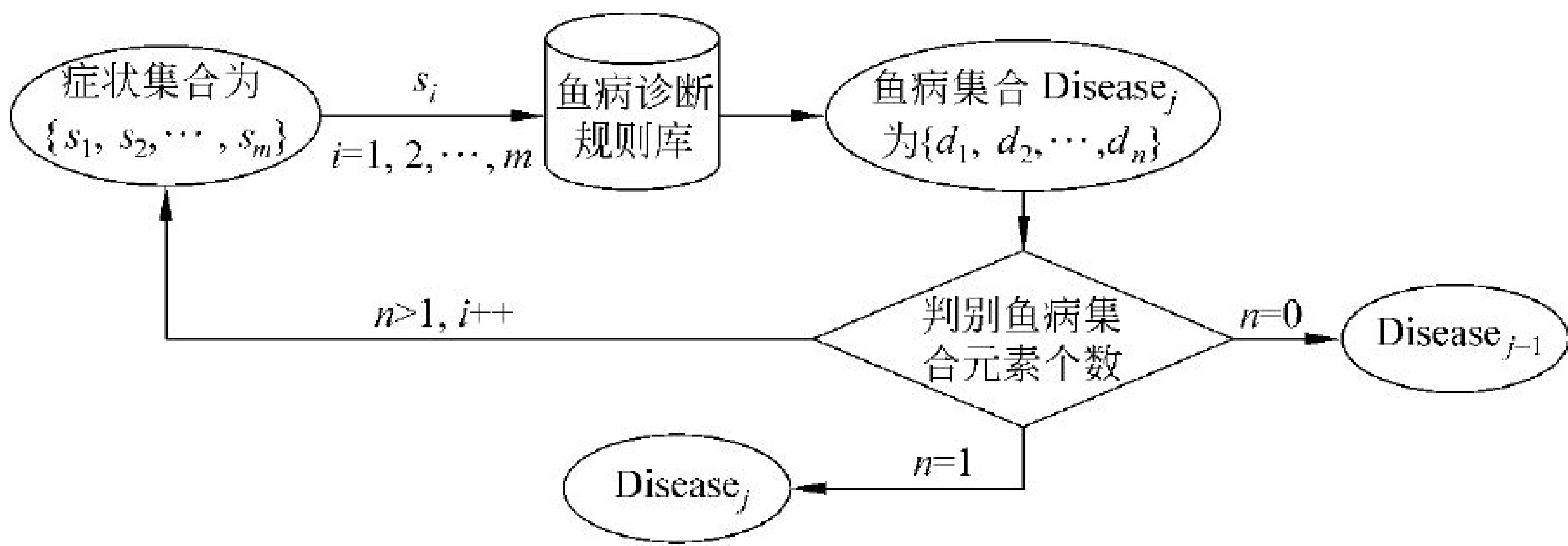


图 7-4　鱼病案例的初级匹配

7.2.3　鱼病案例相似性度量

用户提交的鱼病案例描述经过鱼病症状的特征辨识和初级匹配后被定位到某个子案例库中，下一步的工作就是要从子案例库中检索出最佳案例，即根据一定的相似性度量方法从初级匹配过程中获得的一组候选案例中选取一个或几个与当前问题最相似的案例。案例间的相似性度量是检索的关键，依据某种度量标准对候选案例进行排序，得分最高的就成为最佳案例。

在基于相似性度量的鱼病案例检索模型中，案例检索是根据新问题的描述进行模式匹配的过程。新问题被定义成一组特征属性，检索的目标就是从案例库中找出一个最相似的案例。其检索过程是先根据新问题的初始条件，提取特征属性作为检索目标，再用新问题的特征属性值和案例库中的每个案例对应的特征属性进行相似度量，计算出相似度的大小，找出相似度最大的案例就是最相似的案例。

如前所述，鱼病案例(case)可用一个七元式 CASE＝(DO，M，W，D，S，P，T)来描述。其中，DO 表示对象信息，即什么品种和规格的鱼；M 表示时间信息，即什么季节；W 表示水环境信息，由 pH、溶氧度等指标构成；S 表示症状信息；D 表示疾病信息；P 表示病因信息；T 表示治疗信息。其中最主要的元素分别是：诊断对象，症状，疾病，病因，治疗方案。

案例库中的典型案例，是确诊的案例，用 CASE 来表示。新的鱼病案例，即有待确定疾病、病因和治疗方案的新案例用 $CASE^N$ 来表示。新案例的对象信息、水环境指标、季节、症状信息等是已知的，用四元式 CASE＝(DO，M，W，S)来表示，新案例的问题求解等价于用新问题的已知信息 DO^N，M^N，W^N，S^N 与案例库中各个典型案例的 DO，M，W，S 进行比较，以推理出最相似的典型案例，确定新问题的结论和解决治疗方案 D，P，T。

在鱼病案例知识表示中，以“0”“1”二值表示症状特征属性，用向量表示鱼病案例，这也是为了鱼病案例相似性度量作准备，因为我们对鱼病案例的相似性度量采取最近邻(nearest neighbor)算法，即以距离度量案例之间的相似性。通过计算两个对象在特征空间中的距离来获得两个案例的相似性。在 CBR 系统中，大多数案例检索都使用最近邻算法。

根据鱼病案例的具体特点和诊断要求，设计了鱼病案例最近邻检索模型。

设鱼病案例的特征向量 $\boldsymbol{S}_k=(s_{k1},s_{k2},\cdots,s_{ki},\cdots,s_{kn})$，$s_{ki}$ $(1\leqslant i\leqslant n)$是第 k 个案例的 i 特征属性的症状特征值，$\boldsymbol{S}_k$ 是 n 维鱼病案例特征空间 $\boldsymbol{D}=(D_1,D_2,\cdots,D_n)$上的一点，对于 $\boldsymbol{D}$ 上的案例 $\boldsymbol{S}_k$ 和案例 $\boldsymbol{S}_l$，它们在 $\boldsymbol{D}$ 上的距离为：$\mathrm{Dist}(\boldsymbol{S}_k,\boldsymbol{S}_l)=\left(\sum\limits_{i=1}^{n}D(s_{ki},s_{li})^2\right)^{1/2}$，则案例 $\boldsymbol{S}_k$ 和案例 $\boldsymbol{S}_l$ 的相似度为

$$\mathrm{Sim}(\boldsymbol{S}_k,\boldsymbol{S}_l)=1-\mathrm{Dist}(\boldsymbol{S}_k,\boldsymbol{S}_l)=1-\left(\sum_{i=1}^{n}W_i\times D(s_{ki},s_{li})^2\right)^{1/2}$$

其中，$W_i\in[0,\ 1]$ 是不同症状特征属性的权重，即不同症状特征属性在相似性度量中的重要性。不同的症状对同一疾病的重要性不一样，相同的症状对于不同疾病的重要性也不相同。因此，在度量案例间的相似性时，症状特征属性的权重应根据不同的疾病分别加以确定。根据统计学的观点，如果某特定症状在所有关于某疾病的典型案例中出现的频率很高，那么该症状的出现对确定这种疾病就很重要；反之，频率低则该症状对这种疾病的重要性就低。在此直观地依据这种思想，采用样本统计方法来获得案例库中案例的症状特征属性的权重。

设 s_i 是某种鱼病案例的第 i 个症状特征属性，$P(s_i)$ 为症状属性 s_i 在该鱼病中出现的次数，则症状属性 s_i 在该鱼病中的权重的计算公式是

$$W_i=P(s_i)\Big/\sum_{i=1}^{n}P(s_i)$$

由于对鱼病诊断案例采用了向量(0,1)表示，因此有

$$D(s_{ki},s_{li})=|\ s_{ki}-s_{li}\ |=\begin{cases}0, & s_{ki}=s_{li}\\ 1, & s_{ki}\neq s_{li}\end{cases}$$

当两个案例的特征向量完全相同时，$\mathrm{Dist}(\boldsymbol{S}_k,\boldsymbol{S}_l)=0$，$\mathrm{Sim}(\boldsymbol{S}_k,\boldsymbol{S}_l)=\infty$，即两个案例的相似性为无穷大，完全匹配。一般情况下，有个别特征值不同，可以得出一个相似性值，再根据判别矩阵计算出来的相似度按照大小进行排序，取相似度最大的案例为问题的解。

鱼病案例最近邻检索模型的决策规则如下：

假定有 c 个类别 $w_1,w_2,\cdots,w_c$ 的模式识别问题，每类有标明类别的样本 N_i 个，$i=1,2,\cdots,c$。可以规定 w_i 类的判别函数为：$g_i(x)=\min\limits_i\|x-x_i^k\|$，$k=1,2,\cdots,N_i$。其中 x_i^k 的下角标 i 表示 w_i 类，上角标 k 表示 w_i 类 N_i 样本中的第 k 个。

按照上式，决策规则可以改写为：若 $g_j(x)=\min\limits_i g_i(x)$，$i=1,2,\cdots,c$，则决策 $x\in w_j$。

这一决策方法称为最近邻法。其直观解释是相当简单的，就是说对未知样本 x，只要比较 x 与 $N=\sum\limits_{i=1}^{c}N_i$ 个已知类别的样本之间的距离，并决策 x 与离它最近的样本同类。

在鱼病案例的相似性度量中，对鱼病症状特征值进行了距离计算，因此每个鱼病案例的症状特征值就成了鱼病案例判别的度量指标，由于都是由0或1表示，因此就组成了鱼病案例判别的“0-1”判别矩阵，如表7-3所示。

表 7-3　鱼病案例判别矩阵

鱼病	判别矩阵	鱼病	判别矩阵
肠炎病	$\begin{bmatrix}0&1&1&1&1&0&1&1&1&0&1\\0&0&1&0&0&0&0&1&1&1&0\\0&1&1&0&1&1&0&0&1&0&1\\1&0&0&0&1&1&0&1&1&0&0\\0&1&1&1&1&1&0&1&0&1&1\\1&0&1&0&0&0&0&1&0&0&0\\1&0&0&0&0&0&0&1&0&0&0\\1&0&0&1&1&0&1&1&0&0&0\\1&0&0&0&0&0&0&1&0&0&0\end{bmatrix}$	水质不良所致疾病	$\begin{bmatrix}1&1&1&1&1&0&1&1&0&1&1\\1&0&1&0&1&0&0&1&1&1&0\\1&0&1&0&1&1&0&0&1&0&1\\1&0&0&0&1&1&0&1&1&0&0\\1&1&1&1&0&1&0&1&0&1&1\\1&0&1&0&1&0&1&1&0&1&0\\1&0&0&0&0&0&0&1&0&1&0\\1&0&0&1&1&0&1&1&0&0&0\\1&0&1&0&1&0&0&1&0&1&0\end{bmatrix}$



最近邻法有方法简单的优点，其错误率为：$P^* \leqslant P \leqslant P^*\left(2-\frac{c}{c-1}P^*\right)$，其中 P^* 为贝叶斯错误率，c 为类数。由于 P^* 一般较小，若忽略上式右边括号中第二项，则可粗略表示为 $P^* \leqslant P \leqslant 2P^*$，即近邻法错误率在贝叶斯错误率 P^* 和两倍贝叶斯错误率 $2P^*$ 之间。

7.2.4 鱼病案例复用

一旦检索到相似案例，系统就可以使用检索到的相似案例所提供的解决方案来解决新的问题，这个过程称为案例重用。检索到的源案例中含有对过去问题的解答，把该解答复用于当前的目标案例或新问题的求解，即案例的复用。[3~5]

在建立起个体间的一一对应，并且已知相似性关系后，就可以进行类比转换了。类比转换的目的是要把源案例的全部(或部分)求解方案转换成对目标案例可用的求解方案，并有可能对转换的求解方案进行修改。在简单情况下，只需进行两项工作：将源案例的描述转换成语义相似的目标案例的描述，把中间状态从源案例中复制到目标案例中即可；将源案例的个体转换成目标案例中的与其有一一对应关系的个体。经过上述复制，便可以得到目标案例的一个完整的(或部分的)求解方案。[6,7]

在 CBR 系统中，当案例库中不存在和新案例完全匹配的旧案例时，只能找到一个和待求问题比较相似的旧案例，然后再进行修正，使其能适应新情况，从而得以求解。修正技术可以简单理解为用其他内容替换解决方案的一部分，或者修改整个解决方案。所谓修正可以直接在解决方案中插入一些新内容，也可以删除、替换某一部分内容。我们利用鱼病理论知识和案例知识来学习修正规则。

7.2.5 鱼病案例保存

新问题得到了解决，则形成了一个可能用于将来情形与之相似的问题。这时有必要把它加入到案例库中。在案例保存时要进行模式归纳(schema induction)，即如果类比导致了一个解，则执行两个相似体的泛化过程，以便形成一个抽象的案例模式。换言之，就是寻找两个相似体的共性，然后加以抽象和泛化。在这个过程中，可以发现一般性知识，而且可以将一般性知识存储在知识库中。因此，抽象有助于建立新的语义记忆单元，对知识库进行更为合理的组织。其次，同抽象的案例进行类比，在检索过程中，可以一定程度地减少备选集合中源案例的数量，也可以减少类比过程中其他部分的工作量。[8]

设源案例库 $U=(X_1,X_2,\cdots,X_i,\cdots,X_m)$，$X_i$ 为某源案例，目标案例 Y 与 X_i 的相似度为 $\varepsilon_i\in(0,1)$，η 为专家定义的一个阈值(此处设为 0.5)，案例的学习策略分为以下几种情况：

(1) $\forall\varepsilon_i=0, i\in[1,m]$，新案例与案例库中的所有案例都不匹配，这时新案例可以加入案例库。

(2) $\exists\varepsilon_i=1, i\in[1,m]$，新案例与某源案例完全相似，则新案例不加入案例库。

(3) $\forall\varepsilon_i<\eta, i\in[1,m]$，这时新案例可以加入案例库。

(4) $\exists\varepsilon_i>\eta, i\in[1,m]$，把具有 $\max(\varepsilon_i)$ 的源案例的解决方案改写为新案例的解决方案。

7.3 实例分析

前面针对 CBR 在鱼病诊断推理中的应用进行了详细分析与阐述，下面将对一个具体案例的 CBR 诊断推理过程进行案例分析，包括特征辨识、初级匹配、相似性度量、案例复用与案例保存等各个方面的具体推理过程。

1. 特征辨识

养鱼户通过电话向坐席人员描述病鱼症状：几乎不摄食，逐渐有死亡个体，有的腹部肿胀，肛门红肿，有黄色液体流出，肠道充血，无食物，有脓液，部分有腹水。根据这些症状描述，坐席人员通过系统的特征辨识提示功能来提取出新案例的症状特征属性（厌食，有死亡，腹部肿胀，肛门红肿，肛门有黄色液体流出，肠道充血，肠道无食，肠道有脓液），并通过系统的症状输入模块将这些症状特征属性一一输入到系统中，等待初级匹配。

2. 初级匹配

症状表现向量为 $\boldsymbol{S}_8=(s_1,s_2,s_3,s_4,s_5,s_6,s_7,s_8)$=（厌食，有死亡，腹部肿胀，肛门红肿，肛门有黄色液体流出，肠道充血，肠道无食，肠道有脓液），由它们组合成鱼病-症状矩阵为 **S_D**＝|Symptom⁻，Disease⁻|，进行初级匹配：

$$\text{Select Disease}^- \text{ from } \mathbf{S_D} \text{ where Symptom}^- = S_x, \quad x=1,2,\cdots,8$$

检索出目标 Disease⁻ 是肠炎病，进入针对每个具体案例的相似性度量的下一步。

3. 相似性度量

肠炎病的判别矩阵为

$$\boldsymbol{T}=\begin{bmatrix} 0&1&1&1&1&0&1&1&1&0&1\\ 0&0&1&0&0&0&0&1&1&1&0\\ 0&1&1&0&1&1&0&0&1&0&1\\ 1&0&0&0&1&1&0&1&1&0&0\\ 0&1&1&1&1&1&0&1&0&1&1\\ 1&0&1&0&0&0&1&0&0&0&0\\ 1&0&0&0&0&0&0&1&0&0&0\\ 1&0&0&1&1&0&1&1&0&0&0\\ 1&0&0&0&0&0&0&1&0&0&0 \end{bmatrix}$$

通过与肠炎病的案例特征表进行比照，将具有的症状表示为 1，不具有的症状表示为 0，则此案例的特征向量 $\boldsymbol{S}$ 可以表示为(1,0,1,1,1,0,1,1,0,1,1)，鱼病案例相似性度量公式为

$$\text{Sim}(\boldsymbol{T}_k,\boldsymbol{S})=1-\text{Dis}(\boldsymbol{T}_k,\boldsymbol{S})=1-\left[\sum_{i=1}^{11}W_i\times D(t_{ki}-s_i)^2\right]^{1/2}$$

这里，$\boldsymbol{T}_k$ 是判别矩阵中的第 k 个源案例；t_{ki} 表示第 k 个源案例中第 i 个症状特征属性的值；s_i 表示此案例中第 i 个症状的属性值。

$$D(\boldsymbol{T}_k,\boldsymbol{S})=|t_{ki}-s_i|=\begin{cases}0, & t_{ki}=s_i\\ 1, & t_{ki}\neq s_i\end{cases}$$



症状特征属性的权重：$W=\{w_1,w_2,\cdots,w_{11}\}=\{5,3,5,3,5,3,2,8,4,2,3\}$，$\sum_{i=1}^{11}w_i=43$，

$\{w_i\}=\{0.116,0.652,0.116,0.652,0.116,0.652,0.465,0.186,0.093,0.465,0.652\}$，$i=1,2,\cdots,11$。

$\mathrm{Sim}(\boldsymbol{T}_k,\boldsymbol{S})=(0.875,0.625,0.656,0.594,0.90625,0.728,0.781,0.688,0.742)$。

由上可知，第 5 个源案例与目标案例最相近。

4. 案例复用

第 5 个源案例的症状特征属性为：（厌食，腹部肿胀，肛门红肿，肛门有黄色液体流出，肠道充血）；病因分析：（连天阴雨造成饵料霉变后投喂诱发）；防治措施为：（内服呋喃唑酮、大蒜素、氟派酸 5～7 天，外用二氧化氯全泼）。因此，将这条案例的病因分析和防治措施复用到目标案例。

5. 案例保存

这一步需要向养鱼户进行验证，即先将目标案例和复用的病因分析、防治措施存放到临时库，待养鱼户按照系统得出的防治措施进行诊治后，如果成功，就将其存入源案例库；如果失败，则将其存入失败案例库。

7.4　小结

本章介绍了基于 CBR 和 RBR 协同推理的鱼病诊断推理模型，包括鱼病案例的两级索引模型和基于产生式规则和最近邻算法的案例检索模型。研究结果表明，鱼病案例两级索引模型符合鱼病案例的特点，提高了鱼病案例的检索效率；CBR 与 RBR 协同推理的鱼病诊断推理模型较以往系统有更高的准确性和效率，所采用的产生式规则和最近邻的方法简单实用、便于维护。

参考文献

1. 郭艳红，邓贵仕. 基于事例的推理(CBR)研究综述. 计算机工程与应用，2004，21：1～5
2. 李茹，任海涛，刘开瑛，梁吉业. 基于案例的推理在农业专家系统中的应用. 计算机工程与应用，2004，25：196～199
3. 杜一，张沛超，郁惟墉. 基于事例和规则混合推理的变电站故障诊断系统. 电网技术，2004，1：34～37
4. 段军，耿瑞平，涂序彦. 基于 Rough Set 和神经网络的 CBR 快捷检索方法. 计算机工程与应用，2003，3：25～27
5. 杭小树，熊范伦. 基于 CBR 的农作物病虫害预报专家系统. 计算机工程与应用，2000，10：161～163
6. Nirmalie Wiratunga，Susan Craw，Bruce Taylor，Genevieve Davis. Case-based reasoning for matching SMARTHOUSE technology to people's needs. Knowledge-Based Systems，2004，17：139～146.
7. Pablo Gervás，Belén Díaz-Agudo，Federico Peinado，Raquel Hervás. Story plot generation based on CBR. Knowledge-Based Systems，2005，18：235～242.
8. Petra Perner. Are case-based reasoning and dissimilarity-based classification two sides of the same coin. Engineering Applications of Artificial Intelligence，2002，15：193～203

第 8 章

基于相对熵的 FDT 群决策偏好集结模型

在鱼病实际会诊过程中,会诊专家组中各成员的知识、经验和经历各异,使得各专家的决策偏好存在较大的差异,如何有效地集结群中各成员的决策偏好以形成群体的决策偏好就成为最终确诊决策的一个关键问题。本章对群决策的偏好表示、一致化以及集结模型和方法进行了详细的讨论分析,针对鱼病远程会诊群决策问题的特点,选用了相对熵最优化集结模型构建会诊专家组群决策偏好表示及集结模型,算例结果表明模型是适宜合理的。

8.1 决策偏好的表示方法

作为价值观的表现形式,偏好是人对客观事物的一种价值度量,表现为对事物的喜好程度或接受程度,或对多个事物的不同认可程度的倾向性。

对决策而言,决策者的行为都建立在决策人的偏好结构之上,所谓偏好结构,是指决策人对多个不同的事物(目标、方案、结果等)的不同好恶程度,它反映了决策人的价值取向。偏好,在决策问题中具体表现为对多个方案和目标选取上的一种优先次序关系。因而将这种优先次序关系称为偏好关系或偏好序。

在进行决策分析时,存在如何描述或表达后果对决策人的实际价值,以便反映决策人心目中对各种后果的偏好次序(preference order)的问题。偏好次序是决策人的个性与价值观的反映,与决策人所处的社会地位、经济地位、文化素养、心理和生理(身体)状态有关。

在决策理论中,后果对决策人的实际价值,即决策人对后果的偏好次序是用效用(utility)来描述的。效用就是偏好的量化,是数(实值函数)。200 多年前,Bernoulli 通过对圣彼得堡悖论的研究,提出了用期望偏好(即期望效用)取代期望收益评价决策问题中各种行动的优劣。Von Neumann-Morgenstern 则对此进行公理化解释:在给定的公理体系下,存在一个在正线性变换下唯一的效用函数,使决策问题的求解准则:选择某个行动,使期望效用最大化。这一准则称为 Bernoulli 规则。按照 Von Neumann-Morgenstern 的公理体系所设定的效用函数是基数性的。20 世纪 50 年代,Arrow 在讨论社会选择时指出,进行效用的人际比较时,效用的基数性毫无意义,因此 Arrow 只讨论序数效用函数。

基数效用函数与序数效用函数的区别在于:

(1) 基数效用函数定义在展望集 P 上(考虑后果及其概率分布),是实数;序数效用函数可以定义在后果集 C 上,不涉及概率,可以是正整数。

(2) 基数效用函数不仅反映决策人对后果的偏好次序,还能反映决策人的偏好强度,在正线性变换下唯一;若 2,2.01,100 是某三个行动的展望的基数效用函数,则变换为 $2b+c$,$2.01b+c$,$100b+c$(其中 $b、c\in \mathbf{R}^1$,$b>0$)之后仍是这三个行动的展望的效用函数。而序数效用



并不反映偏好强度，只反映决策人对后果的偏好次序，它在保序变换下唯一；原第一、第二和第三作为序数信息，若变换成数值越大越优的序数效用函数，则可变为9，4，1或6，4，2或4，2，1，等等。

与Arrow分享1972年诺贝尔经济奖的Hicks对效用函数的基数性和序数性作过一个十分形象的比喻：如果知道两个人的身高，那么我可以把高个儿排在第一位，但是不知道身高也没关系，让他们两个人比一下不也能把高个儿找出来吗？

传统选择理论认为，每个人从商品的消费中获得的满足可以用数值来衡量。Hicks认为这一点并不必要，人们并不是非得知道自己能够获得多大的满足才能进行选择。作为对新古典经济理论的发展，Hicks用序数效用理论重新表述了价值理论，这对于经济学的选择理论而言意义重大。

本章对群决策中常见的不同偏好信息的形式、一致化方法以及集结模型和方法进行了综述分类，并且针对鱼病会诊决策的研究现状，对不同方法和模型的适用性进行了对比分析，以期找到合适的模型和方法来求解鱼病远程会诊群决策的偏好集结问题。其中包括很多公式，为了便于理解和使用这些公式，给出群决策偏好集结的公式关系框图，如图8-1所示。在实际使用中，可以根据不同的偏好形式选择不同的模型和方法。

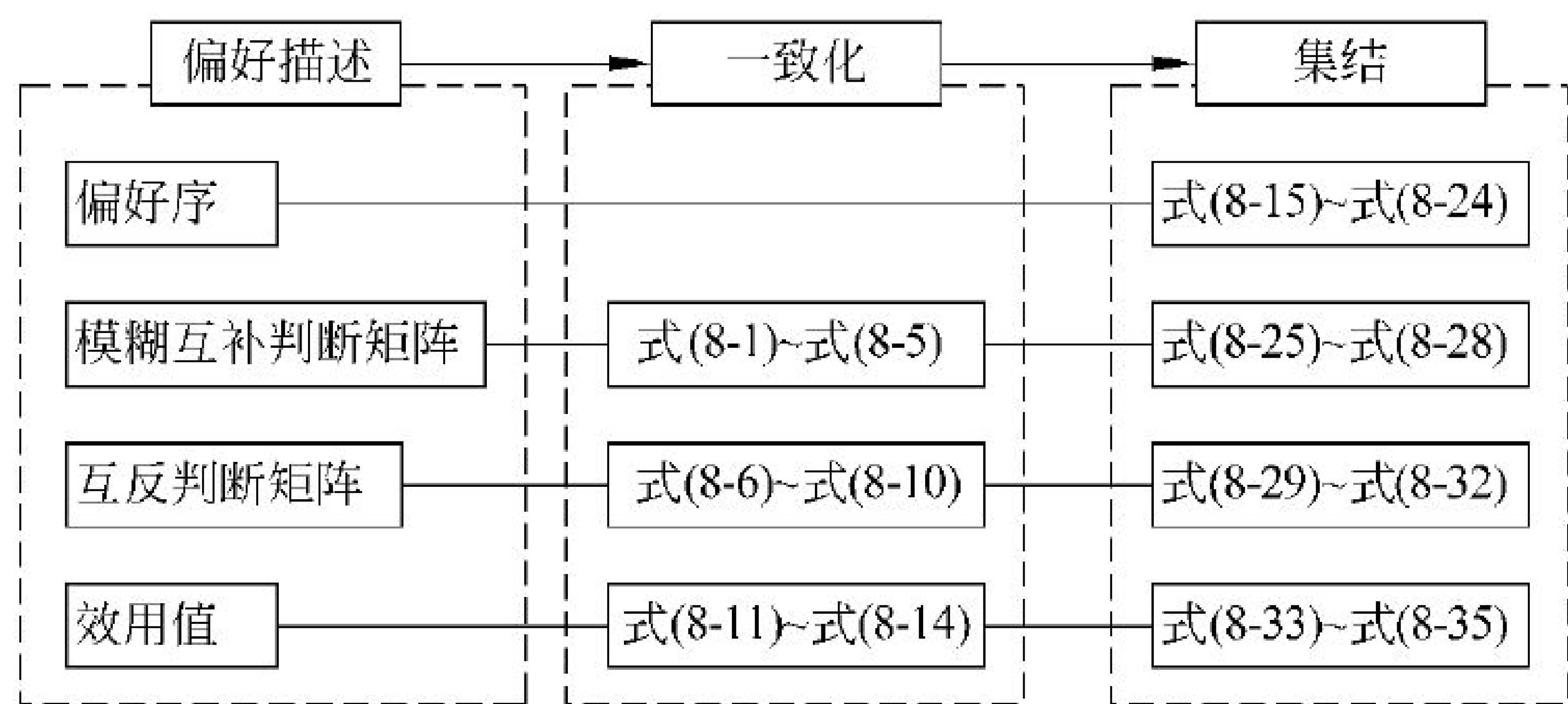


图8-1 公式关系框图

8.2 常见偏好信息的数学描述

在现实中，许多决策问题往往不是由个别决策者作出，而是由多个相关决策者(专家群体)共同制订的。由于决策个体之间在经验、知识以及事物看法上的差异性，再加上决策对象本身的模糊性和不确定性，在实际决策过程中，即使针对同一问题，决策者也完全有可能给出不同形式的偏好信息。因此，有关多种形式偏好信息并存的群决策问题近年来已引起了重视。

目前已见到的主要研究成果有：Chiclana F，Herrera F等在国际上首次提出了研究具有不同形式偏好信息的群决策问题的重要性，并给出了效用值、序关系值和模糊互补判断矩阵等3种形式偏好信息的集结方法。[1]徐泽水、樊治平等和宋光兴等研究了具有模糊判断矩阵和AHP(analytil hierarchy process)判断矩阵两种形式偏好信息的群决策方法。[2~4]Herrera F，Herrera Viedma E等研究了偏好序和效用函数转化为AHP判断矩阵的方法。[5]Delgado M，Herrera F等给出了群决策中具有语言判断矩阵和数值判断矩阵两种形式偏好信息的集结方法。[6]在此基础上，徐泽水和吴江研究了群决策中常见的四类偏好信息，即偏好次序型、效用值型、互补判断矩阵型和互反判断矩阵型。[7,8]王欣荣，樊治平则增加了区间数评价值和模糊语

言评价值两种偏好信息形式，研究了 6 种不同偏好信息情况下的群决策问题。[9]

如前所述，我们研究的鱼病远程会诊群决策问题是给出方案偏好信息的多属性群决策问题。通常，这类群决策中决策者可能给出的偏好信息有两大类，一类是直接给出方案优劣的排序，如偏好序关系值，效用值（或评价值）等；另一类是给出方案优劣的间接偏好，如方案两两比较判断矩阵的形式。而从评价元素的表示方式来看，有 4 种类型，分别为实数值、区间数、模糊数以及自然语言符号。一般来说，信息不完全时，可采用序关系值表示；而信息较充分时可用基数效用值；当信息模糊时，可采用模糊数或灰效用来表示。

近年来，由于决策者使用自然语言对方案进行判断具有容易表达和易于作出决策的特点，判断更接近决策者的思维习惯，基于自然语言评价信息的群决策成为研究热点。而对这类模糊偏好信息，常常转化为区间数、三角模糊数、梯形模糊数或直觉模糊集等效用值的形式，然后依据模糊集理论对其进行研究。

在实际应用中，效用值、序关系值、模糊互补判断矩阵和互反判断矩阵是经常用到的四种偏好信息形式。在鱼病会诊决策中，基本是定性决策，很少使用定量决策，信息获取比较困难，因此为了便于研究，评价元素初步以实数型为主。因此，这里只给出这四种常见偏好信息的形式。

设群的成员有 m 个，所有的决策成员 $d_k(k=1,2,\cdots,m)$ 构成决策群体 $D=\{d_1,d_2,\cdots,d_m\}$。针对有限备选方案集 $X=\{x_1,x_2,\cdots,x_n\}$（$n\geqslant 2$，其中 x_i 表示第 i 个决策方案），决策成员 d_k 可能给出的偏好信息在形式上通常有如下几种：

（1）偏好序关系值

对于方案集 X，决策者 d_k 按自己的偏好给出次序关系向量：

$$\boldsymbol{O}^k=(o_1^k,o_2^k,\cdots,o_n^k)^{\mathrm{T}}$$

其中，$o_i^k\in\mathbb{N}$，表示方案 x_i 在所有方案中的位置次序，o_i^k 越小，对应的方案 x_i 越优。这是一个整数型决策向量。

偏好序本身有着天然的优点，与其他方法相比，它对决策个体所要求显现的信息最少，只需对方案进行比较即可，这使得它在实际中非常容易操作而且可以在需要输入信息非常少的许多领域中应用。因此在鱼病远程会诊群决策的初期研究阶段，偏好序是会诊专家群体最容易给出的偏好信息形式之一。

（2）效用值（或评价值）

对于方案集 X，决策者 d_k 按自己的偏好给出效用值向量如下：

$$\boldsymbol{U}^k=(u_1^k,u_2^k,\cdots,u_n^k)^{\mathrm{T}}$$

其中，$u_i^k\in(0,1)$，是一个 0 到 1 之间的实数型数值，表示决策者 d_k 给出方案 x_i 的基数效用值，效用值 u_i^k 越大，对应的方案 x_i 越优。因此这里没有特别说明的效用值都是基数效用值。

（3）模糊互补判断矩阵

设决策者 d_k 对 X 中方案进行两两比较后给出的互补判断矩阵为 $\boldsymbol{B}^k=(b_{ij}^k)_{n\times n}$，其中，$b_{ij}^k$ 表示决策者 d_k 认为方案 x_i 优于方案 $x_j(x_i\succ x_j)$ 的程度，$b_{ij}^k\geqslant 0$，$b_{ij}^k+b_{ji}^k=1$，$b_{ii}^k=0.5$，$\forall\, i,j,k\in\mathbf{N}$。

构造模糊互补判断矩阵时，可以采用的标度有 0～1 三标度、0～2 三标度、−2～2 五标度、0.1～0.9 五标度以及 0.1～0.9 九标度。常用的三种互补标度见表 8-1。在实际鱼病会诊决策中，需要排序的一组方案往往已经过初步筛选，一个方案一般不会极端优于另一个方案。因此，建议会诊专家在构造模糊互补判断矩阵时使用 0.1～0.9 标度。如无特殊说明，以下讨论



中所涉及的模糊互补判断矩阵均建立在0.1～0.9五标度的基础上。

表8-1　三种互补标度

0～1三标度	0.1～0.9五标度	0.1～0.9九标度及含义	
0	0.1	0.100	元素 i 极端重要于元素 j
		0.138	元素 i 强烈重要于元素 j
	0.3	0.325	元素 i 明显重要于元素 j
		0.439	元素 i 稍微重要于元素 j
0.5	0.5	0.500	元素 i 与元素 j 同样重要
		0.561	元素 i 稍微重要于元素 j
	0.7	0.675	元素 i 明显重要于元素 j
		0.862	元素 i 强烈重要于元素 j
1	0.9	0.900	元素 i 极端重要于元素 j

资料来源：徐泽水.不确定多属性决策方法及应用[M].北京：清华大学出版社，2004，28

（4）互反判断矩阵

设决策者 d_k 对 X 中方案进行两两比较后，给出的互反判断矩阵为 $\mathbf{A}^k=(a_{ij}^k)_{n\times n}$，其中，$a_{ij}^k$ 是由决策者依据Satty提出的1～9标度法给出，表示决策者 d_k 认为方案 x_i 相对于方案 x_j 的重要程度，且有 $1/9\leqslant a_{ij}^k\leqslant 9$，$a_{ii}^k=1$，$a_{ij}^k a_{ji}^k=1$，$\forall\, i,j,k\in\mathbf{N}$。

构造互反判断矩阵时，可采用的标度主要有1～9五标度、指数五标度、9/9～9/1五标度以及10/10～18/2标度，如表8-2所示。实际决策中，以Satty提出的1～9五标度法最为常用。因此，这里所涉及的互反判断矩阵均建立在此标度法的基础上。

表8-2　四种互反标度

1～9标度	指数标度	10/10～18/2标度	9/9～9/1标度	含义
1	a^0	10/10	9/9	元素 i 与元素 j 同样重要
3	a^2	12/8	9/7	元素 i 稍微重要于元素 j
5	a^4	14/6	9/5	元素 i 明显重要于元素 j
7	a^6	16/4	9/3	元素 i 强烈重要于元素 j
9	a^8	18/2	9/1	元素 i 极端重要于元素 j

资料来源：徐泽水.不确定多属性决策方法及应用[M].北京：清华大学出版社，2004，28

8.3　群决策中不同形式偏好信息的一致化研究

在群决策中，由于各决策者的差异性，以及决策对象的复杂性，不同决策者对同一决策问题即使在同一时空也可能给出不同形式的偏好信息。群决策方法有两种，一种方法是先对各种偏好信息分别排序，再对序关系进行集结；另一种方法是先把各类不同的偏好信息转化为同类偏好信息，再对同类偏好信息进行集结排序。一般来说，后一种方法更能体现群体的统一意志。

目前，对各类偏好信息的转换关系研究已经受到越来越多人的关注。但在转化过程中，如果将不同形式的偏好信息均转化为序关系值后再集结，会丢失较多的决策信息。因此目前的转换方法多是将不同偏好信息转换为互补判断矩阵或者互反判断矩阵，也有将不同偏好信息转换为效用值。例如，肖四汉等研究了群决策中具有互反判断矩阵和模糊互补判断矩阵两种

形式偏好信息的转化方法。Herrera F 等、徐泽水和樊治平等人研究了效用值、序关系值和模糊互补判断矩阵转化为互反判断矩阵的偏好信息的集结方法。肖四汉等研究多个决策者给出关于 Fuzzy 判断矩阵和 AHP 判断矩阵，建立了能够集成这两类偏好信息的最优化模型。樊治平等给出将不同形式的偏好信息转化为模糊判断矩阵形式的计算公式，然后基于 OWA 算子给出集结各决策者偏好信息和方案优选的方法。吴江分别利用差值和比值关系给出了 4 种偏好信息与判断矩阵的转换方法。陈华友给出互反判断矩阵、模糊互补判断矩阵和序关系值转化为效用值的一种方法。[10~16]

本章的研究思路是将群决策理论与方法引入鱼病远程会诊领域，因此在对最新研究进展加以分析的基础上，选择适合鱼病远程会诊群决策问题的方法。下面介绍不同偏好信息的转化方法。

8.3.1 不同偏好信息转换为判断矩阵的一致化方法

假设：一般地，决策者 d_k 认为 n 个备选方案中排序列第 1 位的方案 x_i 的效用值为 1，排最后 1 位(第 n 位)的方案 x_j 的效用值为 0；方案 x_i 优于方案 x_j 的程度为 1，方案 x_i 对方案 x_j 的相对重要程度为 9；两相同方案中任一方案优于另一方案的程度为 0.5，相对重要程度为 1。

1. 偏好次序与互补判断矩阵的转换关系

(1)考虑 $b_{ij}^k=f(o_i^k/o_j^k)$。函数 f 为两方案序变量比值(o_i^k/o_j^k)的单调下降函数，给出它们之间一种形如 $f(x)=(x-n)/(a+b(x-n))$ 的双曲线函数关系：

$$b_{ij}^k=\frac{n-o_i^k/o_j^k}{(n-1)(1+o_i^k/o_j^k)}=\frac{n\cdot o_j^k-o_i^k}{(n-1)(o_j^k+o_i^k)} \tag{8-1}$$

(2) 考虑 $b_{ij}^k=f(o_i^k-o_j^k)$。函数 f 为两方案序变量比值$(o_i^k-o_j^k)$的单调下降函数，给出它们之间一种形如 $f(x)=kx+c$ 的线性函数关系式：

$$b_{ij}^k=\frac{1}{2}\left(1-\frac{o_i^k-o_j^k}{b-1}\right) \tag{8-2}$$

2. 效用值与互补判断矩阵的转换关系

(1) 考虑 $b_{ij}^k=f(u_i^k/u_j^k)$。函数 f 为两方案序变量比值(u_i^k/u_j^k)的单调增函数，给出它们之间一种形如 $f(x)=x/(a+bx)$ 的双曲线函数关系：

$$b_{ij}^k=\frac{u_i^k}{u_i^k+u_j^k} \tag{8-3}$$

(2) 考虑 $b_{ij}^k=f(u_i^k-u_j^k)$。函数 f 为两方案序变量比值$(u_i^k-u_j^k)$的单调增函数，给出它们之间一种形如 $f(x)=kx+c$ 的线性函数关系式：

$$b_{ij}^k=\frac{1}{2}(1+(u_i^k-u_j^k)) \tag{8-4}$$

3. 互补判断矩阵与互反判断矩阵的转换关系

徐泽水研究的互补判断矩阵与互反判断矩阵转换关系，即[0.1,0.9]↔[1/9,9]的映射。下面是吴江研究的[0,1]↔[1/9,9]的映射关系。[8]

(1) 假设 $b_{ij}^k=f(a_{ij}^k)$，函数 $f(a_{ij}^k)$为 a_{ij}^k 的单调增函数，考虑形如 $f(x)=c+\log_{81}kx$ 的对数函数关系，即式(8-5)的反函数如下：



$$b_{ij}^k = \frac{1}{2} + \log_{81} a_{ij}^k \tag{8-5}$$

(2) 假设 $a_{ij}^k = f(b_{ij}^k)$，函数 $f(b_{ij}^k)$ 为 b_{ij}^k 的单调增函数，考虑形如 $f(x) = 9^{(ax+b)}$ 的指数函数关系如下：

$$a_{ij}^k = 9^{(2b_{ij}^k - 1)} \tag{8-6}$$

4. 偏好次序与互反判断矩阵的转换关系

(1) 由式(8-1)和式(8-5)得其更为合理的相互关系为

$$a_{ij}^k = 9^{\frac{(n+1)(o_i^k - o_j^k)}{(n-1)(o_i^k + o_j^k)}} \tag{8-7}$$

(2) 由式(8-2)和式(8-6)得其相互关系为：

$$a_{ij}^k = 9^{\frac{o_i^k - o_j^k}{n-1}} \tag{8-8}$$

5. 效用值与互反判断矩阵的转换关系

(1) 由式(8-3)和式(8-5)得其相互关系为

$$a_{ij}^k = \frac{u_i^k - u_j^k}{9^{(u_i^k + u_j^k)}} \tag{8-9}$$

(2) 由式(8-4)和式(8-6)得其相互关系为

$$a_{ij}^k = 9^{u_i^k - u_j^k} \tag{8-10}$$

8.3.2 不同偏好信息转换为效用值的一致化方法

1. 互反判断矩阵转化为效用值的方法

决策者 d_k 给出对方案集 X 中方案进行两两比较的互反判断矩阵 $\boldsymbol{B}^k$，可按特征根法计算 $\boldsymbol{B}^k$ 的权向量 $\boldsymbol{W} = (w_1^k, w_2^k, \cdots, w_n^k)^{\mathrm{T}}$，它实际上是 $\boldsymbol{B}^k$ 的最大特征根 $\lambda_{\max}$ 所对应的归一化特征向量。权向量体现了决策者 d_k 按个人偏好对 X 中各方案重要性程度的认识，因此可以作为各方案的效用值，即

$$u_i^k = w_i^k, \quad i = 1, 2, \cdots, n \tag{8-11}$$

若在精度要求不高的条件下，也可以按方根法计算互反判断矩阵的归一化权向量，其计算公式为

$$u_i^k = w_i^k = \frac{\left|\prod_{j=1}^n b_{ij}^k\right|^{\frac{1}{n}}}{\sum_{i=1}^n \left|\prod_{j=1}^n b_{ij}^k\right|^{\frac{1}{n}}}, \quad i = 1, 2, \cdots, n; k = 1, 2, \cdots, m \tag{8-12}$$

显然由式(8-11)(或式(8-12))的转化，互反判断矩阵的偏好信息一致化为效用值满足 $u_i^k \in (0, 1)$，$\sum_{i=1}^n u_i^k = 1$。

2. 模糊互补判断矩阵转化为效用值的方法

采用模糊互补判断矩阵 $\boldsymbol{P}^k$ 对应的排序向量的一个简捷计算公式，同理可把这个排序向量

$\boldsymbol{W}=(w_1^k,w_2^k,\cdots,w_n^k)^{\mathrm{T}}$ 作为各方案的效用值，即

$$u_i^k=w_i^k=\frac{\sum_{j=1}^{n}p_{ij}^k+\frac{n}{2}-1}{n(n-1)},\quad i=1,2,\cdots,n;k=1,2,\cdots,m \tag{8-13}$$

其中，$p_{ij}^k\geqslant 0$，$p_{ij}^k+p_{ji}^k=1$，$p_{ii}^k=0.5$，$i,j=1,2,\cdots,n$，易证经过式(8-13)的转化，模糊互补判断矩阵的偏好信息一致化为效用值也满足 $u_i^k\in(0,1)$，$\sum_{i=1}^{n}u_i^k=1$。

3. 序关系值转化为效用值的方法

设 o_i^k 为决策者 d_k 给出的第 i 个决策方案 x_i 的位置次序，u_i^k 为其效用值。o_i^k 越小，第 i 个方案 x_i 越优，则其对应的效用值 u_i^k 越大，即 u_i^k 为 o_i^k 的单调下降函数，可以取 u_i^k 为如下形式：

$$u_i^k=\frac{2(n+1-o_i^k)}{n(n+1)},\quad i=1,2,\cdots,n;k=1,2,\cdots,m \tag{8-14}$$

注意到序关系值$\{o_1^k,o_2^k,\cdots,o_n^k\}$是$\{1,2,\cdots,n\}$的一个置换，易证经过式(8-14)的转化，序关系值的偏好信息一致化为效用值也满足 $u_i^k\in(0,1)$，$\sum_{i=1}^{n}u_i^k=1$。

8.4 群决策中不同偏好信息的集结模型和方法

虽然群决策问题的研究起步较早，但由于该问题内在的复杂性，目前在理论和方法上还没有形成比较完整的体系。就其研究目的而言，群决策是为了综合群组中各成员的意见以形成群的意见，并根据群的偏好排列出行动方案的优劣次序。因此，对群组中决策个体偏好信息集结的研究仍是群决策研究的焦点。针对不同形式的决策偏好信息，有相应的集结模型和方法。下面对目前常用的针对不同形式偏好信息的集结模型和方法加以介绍。

8.4.1 个体偏好序的偏好集结函数

个体偏好序集结是建立在二元序关系的基础上。从决策个体对方案集的偏好集结到一致或妥协的群体单一偏好也称为群决策的一致化(consensus)，对个体偏好序进行群体偏好一致化在过去乃至现在一直是社会选择理论研究的一个主题。在个体偏好序的集结研究中，主要是通过在二元比较偏好中建立传递性、连通性、自反性以及对称性等关系，建立任意两个方案的次序关系来显现出决策者对方案的偏好。在对偏好的集结公式研究方面，不同的研究者则采用了各种能反映实际集结关系的数学函数来进行不断的研究。对个体偏好的集结，反映了群决策问题的本质是如何联合个体偏好以形成群体一致或妥协偏好。主要是各种社会选择函数和社会福利函数的研究。

序偏好的一致性集结相对于效用理论而言，所需信息少，每个个体只需表达一个方案对另一个方案的偏好。偏好序的一致性集结在现实中有广泛的应用背景。例如，市场战略设计、表决系统设计、研究开发(R&D)项目优先序排定等。下面是几种常用的社会选择函数。

设 m 个决策者组成决策群体 $D=\{d_1,d_2,\cdots,d_m\}$，方案集为 $A=\{x,y,\cdots\}$。为了方便，方案集有时也记为 $X=\{x_1,x_2,\cdots,x_n\}$。

1. Condorcet 函数

如果存在某个方案 x，能在与其他方案逐一比较时，按过半数规则击败其他方案，则 x 是



Condorcet 方案，群体应选择 x 方案。如果不存在 Condorcet 方案，则按函数

$$f_{c}(x)=\min_{y\in A\setminus\{x\}}\#\{i\mid xP_{i}y\}\tag{8-15}$$

值大小排序方案，其中 $A\setminus\{x\}$ 表示方案集 A 中除去 x 方案后构成的集合。$f_c(x)$ 称为 Condorcet 函数，是 x 与其他所有方案成对比较时得票最少的那一次票数，因此它是极大化极小函数。

2. Borda 函数

每个投票人对 n 个方案进行排序，将 $n-1,n-2,\cdots,1,0$ 这 n 个分数分别赋予优先序为第一位，第二位，……，最后一位的方案。计算各方案的得分总数（称为 Borda 分）。按 Borda 分的高低排列方案的优先序。Borda 函数为

$$f_{B}(x)=\sum_{y\in A}\#\{i\mid xP_{i}y\}\tag{8-16}$$

方案按 $f_B(x)$ 值排序。$f_B(x)$ 是 x 与其他方案逐一比较 $n-1$ 次所得票数之和。

3. Copeland 函数

Copeland 函数为

$$f_{cp}(x)=\#\{y\mid y\in A\wedge xPy\}-\#\{y\mid y\in A\wedge yPx\}\tag{8-17}$$

其中，$\#\{y|y\in A\wedge xPy\}$ 表示 x 按过半数规则能击败 A 中的方案数，$\#\{y|y\in A\wedge yPx\}$ 表示 A 中按过半数规则能击败 x 的方案数，$f_{cp}(x)$ 是 x 与其他方案逐一比较的获胜次数与失败次数之差。

4. Nanson 函数

Nanson 函数是一个 Borda 淘汰过程。每一步得分最低的方案被淘汰，直到无方案可被淘汰为止。可以证明，Nanson 函数符合 Condorcet 原则，即若存在 Condorcet 方案，则它不会被淘汰。

设 $A_1=A$，对 $j\geqslant 1$，令

$$A_{j+1}=A_j\setminus\{x\in A_j;f_B(x)\leqslant f_B(y),\forall y\in A_j\}$$

且
$$\exists y\in A\quad 使\quad f_B(x)<f_B(y)$$

其中，$f_B(x)$ 为 Borda 函数，定义 Nanson 函数为

$$f_{N}(x)=\lim_{j\to\infty}A_j\tag{8-18}$$

5. Dodgson 函数

根据候选方案要能成为简单过半数胜者（或非败者），所需要的改变偏好序的决策者人数来对方案评分。具体计算 $f_D(x)$ 时，可用方案之间成对比较的结果构成 Dodgson 矩阵，并对第 j 行中 $\#\{i|xPy\}<\#\{i|yPx\}$ 的各项求 n_0 与 $\#\{i|xPy\}$ 之差，然后相加得 $f_D(x)$。

$$n_0=\begin{cases}m/2, & m\ 为偶数\\(m+1)/2, & m\ 为奇数\end{cases}$$

设 $n_{jj}=n_0$，则 Dodgson 的社会选择函数为

$$f_{\mathrm{D}}(a_j)=\sum_{k=1}^{m}\frac{|n_0-n_{jk}|+(n_0-n_{jk})}{2} \tag{8-19}$$

$f_{\mathrm{D}}(x)$值越小，方案 x 越优。

6. Kemeny 函数

Kemeny(1972)提出一种偏好集结函数，使一致排序与各决策者对各方案的偏好序有最大的一致性。

定义排序矩阵 $\boldsymbol{L}=[l_{ij}](i,j=1,2,\cdots,n)$，矩阵的元素 l_{ij} 是

$$l_{ij}=\begin{cases}1, & \text{如果 } x_i \text{ 优于 } x_j\\ 0, & \text{如果 } x_i \text{ 与 } x_j \text{ 成平局}\\ -1, & \text{如果 } x_j \text{ 优于 } x_i\end{cases}$$

设 m_{ij} 是认为 x_i 优于 x_j 的决策者人数，m_{ji} 是认为 x_j 优于 x_i 的决策者人数，m_{ij}^* 是认为 x_i 与 x_j 无差异的决策者人数。定义一个比例矩阵 $\boldsymbol{R}$，它的元素 $r_{ij}(i,j=1,2,\cdots,n)$表示一个方案优于另一个方案的决策人数之比。

$$r_{ij}=\frac{m_{ij}+m_{ij}^*}{m},\quad m_{ii}=\frac{1}{2} \tag{8-20}$$

其中，m 是决策者总人数，r_{ij} 表示群中认为 x_iPx_j 的人数在全体成员中所占的比例。

令 $\boldsymbol{G}$ 是一个经过变换得到的投票矩阵，$\boldsymbol{G}=\boldsymbol{R}-\boldsymbol{R}^{\mathrm{T}}$，$\boldsymbol{G}$ 的元素 g_{ij} 表示认为 x_iPx_j 的决策者所占比例与认为 x_jPx_i 的决策者所占比例之差。Kemeny 函数为

$$f_{\mathrm{k}}=\max_{L}\langle G,L\rangle \tag{8-21}$$

其中，$\langle G,L\rangle$表示矩阵 $\boldsymbol{G}$ 与 $\boldsymbol{L}$ 的内积，$\langle G,L\rangle=\sum_{i,j}l_{ij}\cdot g_{ij}$。用$\langle G,L\rangle$的大小来衡量群体偏好的排序与各成员间的一致性。Kemeny 函数是要找 L，使$\langle G,L\rangle$取极大值。

7. Cook-Seiford 函数

Cook 与 Seiford 于 1978 年引入距离函数来度量群体排序与个体排序之间的不一致程度。一致排序定义为使总的绝对距离最小的排序。令 r_{ij} 表示成员 i 把方案 j 排在 r_{ij} 位，$i=1,2,\cdots,m$，$j=1,2,\cdots,n$，即有 m 位委员会成员，有 n 个方案。r_j^{G} 表示群体对方案 j 的排序，成员 i 与群排序的距离为

$$d_i=\sum_{j=1}^{n}|r_{ij}-r_j^{\mathrm{G}}|,\quad i=1,2,\cdots,m \tag{8-22}$$

总距离为

$$d=\sum_{i=1}^{m}d_i=\sum_{i=1}^{m}\sum_{j=1}^{n}|r_{ij}-r_j^{\mathrm{G}}| \tag{8-23}$$

现在要找一种群体排序，使之与各成员排序的总距离最小。由于每个方案只能排在某一位置上，并且每个位置只能挑一个方案，因此可用 0-1 规划求解，即求解式(8-24)。

$$\left.\begin{aligned} &\min \sum_{j=1}^{n}\sum_{k=1}^{n} d_{jk} x_{jk} \\ &\text{s.t}\quad \sum_{j=1}^{n} x_{jk} = 1, \quad k = 1,2,\cdots,n \\ &\qquad \sum_{k=1}^{n} x_{jk} = 1, \quad j = 1,2,\cdots,n \end{aligned}\right\} \tag{8-24}$$

其中若方案 j 排在 k 位，则 $x_{jk}=1$，否则 $x_{jk}=0$。

8.4.2　判断矩阵的偏好集结方法

以上对偏好序的集结函数进行了比较，下面考虑判断矩阵形式的偏好信息的集结。需要指出的是，无论采用哪种集结方法，在集结关于两两比较的偏好信息之前，都有必要检验每个决策者给出的判断信息是否具有一致性，若不具有一致性，则进行集结没有意义。

1. 基于群组互反判断矩阵的偏好信息集结方法

相对于模糊互补判断矩阵而言，互反判断矩阵的排序理论，即偏好集结理论的研究已基本成熟，因此，对群决策中偏好信息的集结多体现在对群组 AHP 排序的研究上，此类研究大体可分为两种类型，一类是判断矩阵元素的加权平均法，另一类是判断矩阵排序向量的加权平均法。概括起来主要有四种方法，即判断矩阵元素的加权几何平均法、加权算术平均法、排序权值的加权几何平均法和加权算术平均法。

为叙述方便起见，设判断矩阵 $\boldsymbol{P}^k=(p_{ij}^k)_{n\times n}$，$k=1,2,\cdots,m$，$\lambda_k$ 为决策者 d_k 的权重，$\lambda_k>0$，$\sum\limits_{k=1}^{m}\lambda_k=1$，其综合排序向量记为 $\boldsymbol{W}=(w_1,w_2,\cdots,w_n)^{\mathrm{T}}$，且满足归一化的约束条件 $\sum\limits_{i=1}^{n}w_i=1$。

（1）用加权几何平均法构造综合判断矩阵 $\boldsymbol{P}^*=(p_{ij})_{n\times n}$

$\boldsymbol{P}^*=(p_{ij})_{n\times n}$ 中，$p_{ij}=\prod\limits_{k=1}^{m}(p_{ij}^k)^{\lambda_k}$，$i,j=1,2,\cdots,n$，并且

$$w_i=\frac{\prod\limits_{j=1}^{n}\sqrt[n]{\prod\limits_{k=1}^{m}(p_{ij}^k)^{\lambda_k}}}{\sum\limits_{i=1}^{n}\prod\limits_{j=1}^{n}\sqrt[n]{\prod\limits_{k=1}^{m}(p_{ij}^k)^{\lambda_k}}},\quad i=1,2,\cdots,n \tag{8-25}$$

（2）用加权算术平均法构造综合判断矩阵 $\boldsymbol{P}^*=(p_{ij})_{n\times n}$

$\boldsymbol{P}^*=(p_{ij})_{n\times n}$ 中，$p_{ij}=\sum\limits_{k=1}^{m}\lambda_k p_{ij}^k$，$i,j=1,2,\cdots,n$，并且

$$w_i=\frac{\sum\limits_{j=1}^{n}\sum\limits_{k=1}^{m}(\lambda_k p_{ij}^k/n)}{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\sum\limits_{k=1}^{m}(\lambda_k p_{ij}^k/n)},\quad i=1,2,\cdots,n \tag{8-26}$$

（3）用加权几何平均法求综合排序权值向量 $\boldsymbol{W}=(w_1,w_2,\cdots,w_n)^{\mathrm{T}}$

在各判断矩阵 $\boldsymbol{P}^k=(p_{ij}^k)_{n\times n}$ 满足一致性的条件下，分别求解对应的等价特征根，得其排序权值向量

$$\boldsymbol{W}^k=(w_1^k,w_2^k,\cdots,w_n^k)^{\mathrm{T}},\quad k=1,2,\cdots,m$$

令 $\tilde{w}_i = \prod_{k=1}^{m}(w_i^k)^{\lambda_k}$，$\boldsymbol{W} = (\tilde{w}_1, \tilde{w}_2, \cdots, \tilde{w}_n)^{\mathrm{T}}$，$i = 1,2,\cdots,n$，则

$$w_i = \frac{\tilde{w}_i}{\sum_{i=1}^{n}\tilde{w}_i} = \frac{\prod_{k=1}^{m}(w_i^k)^{\lambda_k}}{\sum_{i=1}^{n}\prod_{k=1}^{m}(w_i^k)^{\lambda_k}} \tag{8-27}$$

(4) 用加权算术平均法求综合排序权值向量 $\boldsymbol{W}=(w_1, w_2, \cdots, w_n)^{\mathrm{T}}$

在各判断矩阵 $\boldsymbol{P}^k=(p_{ij}^k)_{n\times n}$ 满足一致性的条件下，分别求解对应的等价特征根，得其排序权值向量：

$$\boldsymbol{W}^k = (w_1^k, w_2^k, \cdots, w_n^k)^{\mathrm{T}}, \quad k = 1,2,\cdots,m$$

令 $\tilde{w}_i = \sum_{k=1}^{m}(\lambda_k w_i^k)$，$\boldsymbol{W} = (\tilde{w}_1, \tilde{w}_2, \cdots, \tilde{w}_n)^{\mathrm{T}}$，$i = 1,2,\cdots,n$，则

$$w_i = \frac{\tilde{w}_i}{\sum_{i=1}^{n}\tilde{w}_i} = \frac{\sum_{k=1}^{m}(\lambda_k w_i^k)}{\sum_{i=1}^{n}\sum_{k=1}^{m}(\lambda_k w_i^k)} \tag{8-28}$$

比较上述四种由群成员(决策者)的偏好形成群的偏好的方法，除了上面已经提到的各种方法的明显优缺点外，还有别的一些特性。如从理论上可以证明，当 s 个决策者的判断矩阵均为一致性矩阵时，第一种方法与第三种方法等价。若其中有一些判断矩阵不一致时，两种方法的结果也不完全相同。另外，通过仿真计算可知，当成员数增大时，第三种和第四种方法较第一种和第二种方法有明显优势。因此，再考虑到第一种和第二种方法构造群判断矩阵时在一致性和互反性上所遇到的问题，一般情况下采用第三种方法和第四种方法是适宜的。此外，还有一些基于群组 AHP 的偏好信息集结方法，如王应明提出的最小二乘方法，樊治平采用近年来最新发展的有序几何加权(OWG，ordered weighted geometric)算子对群组偏好信息进行的集结等。[11]

2. 基于群组互补判断矩阵的偏好信息集结方法

对群组互补判断矩阵的集结方法主要有最小方差法，以及有序加权平均(OWA，ordered weighted averaging)算子。由于 OWA 算子在计算上的方便性，近年来在诸多领域得到广泛应用。

定义 8.1 设 $f: R^n \to R$，若 $f(a_1, a_2, \cdots, a_n) = \sum_{j=1}^{n} w_j b_j$，其中 $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^{\mathrm{T}}$ 是与 f 相关联的加权向量，$w_j \in [0,1]$，$\sum_{j=1}^{n} w_j = 1$，且 b_j 是一组数据 $a_l(l=1,2,\cdots,n)$ 中第 j 个最大的元素，则称函数 f 是 n 维有序加权算子，简称为 OWA 算子。

OWA 算子的根本特点是，对数据 $a_l(l=1,2,\cdots,n)$ 按从大到小的顺序重新排序并通过加权集结，而且元素 a_l 与 w_i 没有任何联系，w_i 只与集结过程中顺序的第 l 个位置有关，因此又称为位置向量。

假设有 n 个方案，决策 $d_1, d_2, \cdots, d_k$ 对这 n 个方案给出的不同形式的偏好信息已经一致化为 k 个模糊互补判断矩阵，即 $\boldsymbol{B}^1, \boldsymbol{B}^2, \cdots, \boldsymbol{B}^k$，下面运用 OWA 算子将各决策者的偏好信息集结成群组的偏好，集结步骤如下：



（1）计算群集结的 OWA 算子的权向量 $\boldsymbol{W}=(w_1,w_2,\cdots,w_k)^{\mathrm{T}}$，$w_m\geqslant 0$ 且 $\sum_{m=1}^{k}w_m=1$，其中，$w_m=Q\left(\frac{m}{k}\right)-Q\left(\frac{m-1}{k}\right)$，$\forall k\in M$，式中，$Q(r)$ 为模糊量化算子，由下面的公式给出：

$$Q(r)=\begin{cases}0, & r<\alpha \\ \dfrac{r-\alpha}{\beta-\alpha}, & \alpha\leqslant r\leqslant\beta \\ 1, & r>\beta\end{cases} \tag{8-29}$$

其中，$\alpha,\beta,r\in[0,1]$，在“大多数”、“至少一半”和“尽可能多”的原则下，模糊量化算子 Q 对应的参数(α,β)分别为(0.3,0.8)，(0,0.5)和(0.5,1)。

（2）利用 OWA 算子将各决策者的偏好信息 $\boldsymbol{B}^1,\boldsymbol{B}^2,\cdots,\boldsymbol{B}^k$ 集结为群组的偏好 $\boldsymbol{B}^*=(b_{ij}^*)_{n\times n}$，其中 b_{ij}^* 表示群体认为方案 x_i 优于方案 x_j 的程度，计算公式为

$$b_{ij}^*=\phi_Q(b_{ij}^1,b_{ij}^2,\cdots,b_{ij}^k)=\sum_{m=1}^{k}\omega_m b_{ij}^m,\quad \forall i,j\in\mathbf{N} \tag{8-30}$$

上式中，ϕ_Q 是 OWA 算子；b_{ij}^m 是集合$\{b_{ij}^1,b_{ij}^2,\cdots,b_{ij}^k\}$中按大小排在第 m 位的那个元素。

（3）计算方案优选对应的 OWA 集结算子的权重 $\boldsymbol{w}^*=(w_1^*,w_2^*,\cdots,w_k^*)^{\mathrm{T}}$，其中，$w_q^*\geqslant 0$，$\sum_{q=1}^{n}w_q^*=1$，$w_q^*=Q\left|\frac{q}{n}\right|-Q\left|\frac{q-1}{n}\right|$，$q\in\mathbf{N}$。

（4）在相应原则下(一般选取“大多数”原则)计算方案 x_i 优于其他所有方案的程度 dd_i

$$dd_i=\phi_Q(p_{ij}^*)=\sum_{q=1}^{n-1}w_q^*c_i^q,\quad \forall i,j\in\mathbf{N},j\neq i \tag{8-31}$$

其中，c_i^q 是集合$\{p_{ij}^*\mid j=1,2,\cdots,n,j\neq i\}$中按大小排在第 q 位的元素。

按照计算得到的 dd_i 数值的大小即可对各方案进行排序，如果想进一步得到群组决策者对各方案的主观权重信息，则需要进行第(5)步工作。

（5）计算各方案的权重 λ，即群组对各方案的偏好信息，计算公式如下：

$$\lambda_i=\frac{\sum_{i=1}^{n}dd_i}{n} \tag{8-32}$$

8.4.3　基于效用值的相对熵群决策集结最优化模型

群决策问题本质上是一个集结问题。如果假设决策问题存在着一个客观事实或由决策者群体共同认可的标准，而决策个体给出的方案集合中的方案评估值是相对于客观事实或认可的标准的波动或偏离。在这种假设条件下，群决策问题可以转换为一个优化问题。而熵及其熵优化理论是一种较好的优化工具。

针对这个问题，邱菀华等在对群决策问题实质进行深入分析的基础上，利用信息论中的相对熵概念建立起群决策的相对熵模型，通过对该相对熵模型解的性质合理性分析，认为由相对熵模型得出的几何平均值解具有良好的性质。[19]在此基础上，陈华友和刘春林探讨了一个基于不同偏好信息形式的相对熵最优化集结模型。[16]

下面对相对熵模型的概念和原理进行介绍。

定义 8.2 设 $x_i, y_i \geqslant 0 (i=1,2,\cdots,n)$，且 $1=\sum_{i=1}^{n} x_i \geqslant \sum_{i=1}^{n} y_i$，则称 $h(X,Y)=\sum_{i=1}^{n} x_i \log \frac{x_i}{y_i}$ 为 X 相对于 Y 的相对熵，其中 $X=\{x_1,x_2,\cdots,x_n\}$，$Y=\{y_1,y_2,\cdots,y_n\}$。

定理 8.1 若 $h(X,Y)$ 为 X 相对于 Y 的相对熵，则有

$$h(X,Y)=\sum_{i=1}^{n} x_i \log \frac{x_i}{y_i} \geqslant 0 \tag{8-33}$$

$$\sum_{i=1}^{n} x_i \log \frac{x_i}{y_i}=0, \quad \text{当且仅当 } x_i=y_i; i=1,2,\cdots,n。$$

定理 8.1 表明当两个离散概率分布 X 和 Y 完全相等时，X 相对于 Y 的相对熵达到了最小值，因此可以用相对熵来度量两者的符合程度。

设决策方案集合为 $X=\{x_i\}, i=1,2,\cdots,n,$，决策群体的集合为 $D=\{d_k\}, k=1,2,\cdots,m$，且每个决策者根据自己的偏好分别对决策方案集合 X 中的所有方案给出评判值。我们假设对方案 x_i 的评判值可以用基数效用来表示，并且评判值越大表示决策人对该方案越偏好。集合 $W=\{w_k\}(k=1,2,\cdots,m)$，$\sum_{k}^{m} w_k=1$ 是决策者 d_k 的权重集合，其中 w_k 对应于第 k 个决策者的权重。

形式化表示为：$\forall d_k \in D$，可给出一个映射 $\pi_k: u_i \rightarrow u_i^k$，其中 u_i^k 为决策者 d_k 对方案 x_i 的评判值，为了便于计算，用 u_{ki} 表示。设群体偏好的映射为 $\pi_g: a_i \rightarrow u_{gi}$，群偏好向量为 $\boldsymbol{U}_g=\{u_{gi}\}$，$i=1,2,\cdots,n$。求出每一个 u_{gi} 后，我们就可以根据 u_{gi} 的排序大小对决策方案集 $X=\{x_i\}, i=1,2,\cdots,n$ 进行选择，同时也可以比较两个方案之间的偏好差别。

如果我们将决策者 d_k 对决策方案集合 X 中各方案的偏好评判值假设为对各方案偏好效用的概率测度，每一个决策者对方案集合中所有方案的离散概率测度就形成决策方案集合 X 的一个概率分布。假设专家是在相互独立的情况下对方案集合中各方案进行评判的。因此我们认为作为对各方案偏好效用的概率测度也是相互独立的离散随机变量。现在，建立相对熵集结模型来集结群体的一致或妥协偏好评判值。

根据定理 8.1 知，要想获得一致的群体效用值向量的概率分布，就要最小化每一个决策者给出的所有方案效用值向量的概率分布的相对熵，因此依据上述分析，为得到决策群体对方案集 X 的群偏好向量 $\boldsymbol{U}_g$，须解下面的非线性规划问题：

$$\left.\begin{aligned} &\min Q(\boldsymbol{U}_g)=\sum_{k=1}^{m} w_k \sum_{i=1}^{n}\left(\log u_{gi}-\log \frac{u_{ki}}{\sum_{i=1}^{n} u_{ki}}\right) \cdot u_{gi} \\ &\sum_{i=1}^{n} u_{gi}=1 \\ &u_{gi}>0 \end{aligned}\right\} \tag{8-34}$$

由上式看出，每个决策者对方案集合 X 中的方案的偏好效用值的大小通过规范化限定在区间[0,1]中，通过对各个决策者的偏好效用值与群偏好向量之间偏离值的极小化，使得各个决策者偏好效用与群偏好效用进行了比较，而且对各个决策者之间的偏好效用也进行了比较，最后通过解此非线性规划我们可以得到群决策 REM(minimum relative entropy)算法。

定理 8.2 式(8-34)确定的最优化模型有全局最优解 $\boldsymbol{U}_g^*=(u_{g1}^*, u_{g2}^*, \cdots, u_{gn}^*)^{\mathrm{T}}$，其中



$$u_{gj}^{*} = \frac{\prod_{i=1}^{m}(b_{ki})^{w_k}}{\sum_{i=1}^{n}\prod_{k=1}^{m}(b_{ki})^{w_k}}, \quad i = 1,2,\cdots,n;k = 1,2,\cdots,m \tag{8-35}$$

其中，$b_{ki} = \dfrac{u_{ki}}{\sum_{i=1}^{n} u_{ki}}$，$u_{ki} \neq 0$。

定理 8.2 的详细证明参见文献[6]。在 REM 集结模型中，我们假设式(8-35)中 $u_{ki} \neq 0$，这时 REM 集结规则满足 Arrow 理性条件中的非独裁条件。即决策者对方案的评判值不能为零。如果不是这样，假如某一个 $u_{ki}=0$，则 $b_{ki}=0$，亦即专家 d_k 对方案 x_i 的评估概率估计为 0，这时依据 REM 集结模型专家群体对方案 x_i 的评估概率估计也等于 0。这说明如果专家 d_k 有肯定的判断，不管其他专家意见如何，专家群体都将把专家 d_k 的意见作为自己的判断，这时就出现了独裁现象。

8.5　不同形式偏好集结的 FDT 群决策流程

如前所述，在鱼病诊断领域，基本是依靠鱼病专家或者鱼病专家系统进行定性决策，很少使用定量决策，尤其目前尚未进行会诊决策，信息获取比较困难。因此为了方便研究，在和领域专家充分交流探讨的基础上，偏好信息选用效用值、序关系值、模糊互补判断矩阵和互反判断矩阵这四种经常用到的偏好信息形式，且评价元素初步以实数型为主。同时，在保证有效性的前提下，兼顾实际应用中的可操作性及计算量大小，选用将四种不同形式的偏好信息转化为效用值的一致化方法，用相对熵集结模型来求解，以提高运算效率和决策效果。

下面给出 FDT 群决策中基于不同偏好信息的相对熵方法的算法流程(如图 8-1 所示)：

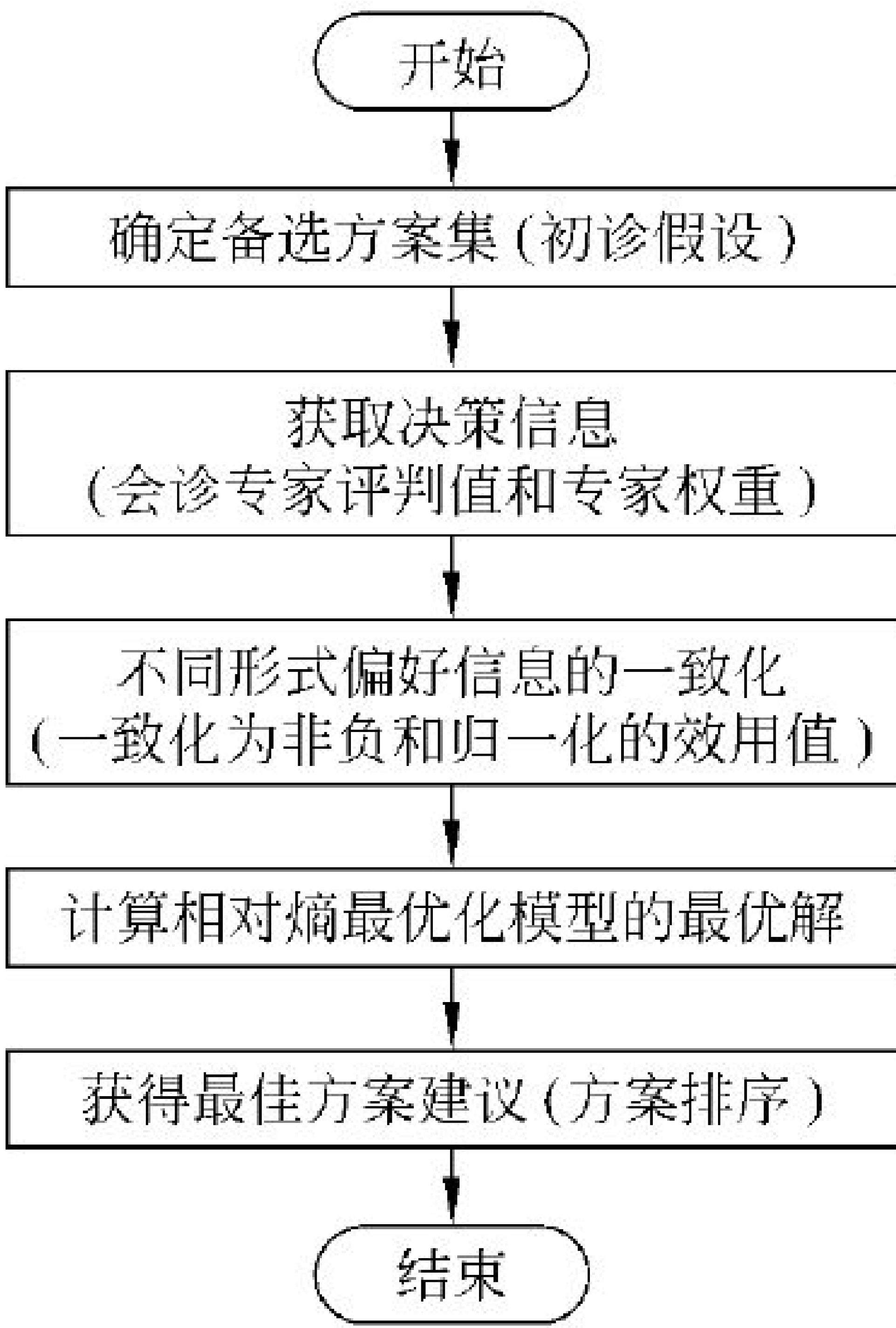


图 8-2　基于相对熵集结模型的 FDT 群决策过程

① 在鱼类复杂病症会诊群决策时，参与会诊的每个鱼病专家决策者在深入分析病症的基础上，可以凭自己的经验和知识或者借助鱼病诊断专家系统构造鱼病的初诊方案(备选方案)集 $X=\{x_1,x_2,\cdots,x_n\}(n\geqslant 2)$，即几种诊断假设；

② 针对初诊方案集，专家们在通过 FDT_GDSS 进行充分讨论和交流的基础上，结合自己的习惯给出对各备选方案的偏好信息(可以是偏好序、效用值、互补或互反判断矩阵中任意一种)，并由会诊组织者给出各会诊专家的权重；

③ 由式(8-11)或式(8-12)、式(8-13)和式(8-14)将会诊专家给出的互反判断矩阵、模糊互补判断矩阵或序关系值的偏好信息转化为非负性和归一化的效用值的偏好信息；

④ 根据式(8-35)计算最优化模型的全局最优解 $\boldsymbol{U}_g^{*}=(u_{g1}^{*},u_{g2}^{*},\cdots,u_{gn}^{*})^{\mathrm{T}}$；

⑤ 根据计算出的最优群体效用值向量 $\boldsymbol{U}_g^{*}$ 中 u_{gi}^{*} 的大小对决策方案集 X 进行排序，从而获得最佳决策方案(即会诊建议)。

8.6　FDT 实例分析

为了说明所选模型的有效性,分别从两个角度进行了验算对比。实际应用时还可以选用别的模型和方法来对比求解,从而保证不同算法的有效性和稳定性。同时,为了使所选案例具有典型性,我们要求专家分别给出4种不同形式的偏好信息,并且假定各会诊决策专家的权重都相等。实际应用时,专家的偏好信息可以依据个人习惯来评判,专家权重可以由会诊组织者依据各专家的经验水平来确定。如果专家给出的偏好信息为同一形式,可以根据上文给出的相应的偏好集结的模型和方法来求解。

例如:某次遇到一个复杂病症需进行会诊,病鱼主要表现为口腔、颌、眼眶、鳃盖及各鳍条充血,体表充血呈斑点状出血,鳃丝肿胀;有些病鱼腹腔积水,肛门红肿。专家组通过鱼病诊断专家系统初步诊断为暴发型出血性败血病,但还是不能确诊具体是哪一种病型,请专家组会诊决策。专家组可以凭经验和知识或者借助鱼病诊断专家系统提出诊断假设集,即本问题的备选方案集 $X=\{x_1,x_2,x_3,x_4\}$={急性甲型,急性乙型,慢性型,潜伏型},并邀请4位专家作为决策群体 $D=\{d_1,d_2,d_3,d_4\}$。各专家依据要求分别对4种方案给出序关系值、效用值、模糊互补判断矩阵和互反判断矩阵4种不同形式的偏好信息如下:

$$d_1:O^1=\{2,1,4,3\}$$

$$d_2:U^1=\{0.4,0.6,0.9,0.1\}$$

$$d_3:\boldsymbol{B}^1=\begin{bmatrix}0.5&0.2&0.1&0.8\\0.8&0.5&0.4&0.6\\0.9&0.6&0.5&0.7\\0.2&0.4&0.3&0.5\end{bmatrix}$$

$$d_4:\boldsymbol{A}^1=\begin{bmatrix}1&1/7&1/3&1/5\\7&1&8&2\\3&1/8&1&1/2\\5&1/2&2&1\end{bmatrix}$$

算法1:假定各会诊专家的权重相等,即 $w_1=w_2=w_3=w_4=1/4$。

① 根据式(8-11)~式(8-14)将4种不同偏好信息一致化为非负和归一化的效用值。

$$U^1=\{0.3000,0.4000,0.1000,0.2000\}$$

$$U^2=\{0.2000,0.3000,0.4500,0.0500\}$$

$$U^3=\{0.2167,0.2750,0.3083,0.2000\}$$

$$U^4=\{0.0546,0.5688,0.1151,0.2615\}$$

② 将计算出的一致化效用值代入式(8-35),计算出最优化模型的全局最优解 $\boldsymbol{U}_g^*$,即:

$$u_1^*=0.1846,\quad u_2^*=0.4185,\quad u_3^*=0.2260,\quad u_4^*=0.1710$$

③ 根据计算出的最优群体效用值的大小对决策方案集 X 进行排序得如下结果:

$$x_2\succ x_3\succ x_1\succ x_4$$

因此最优方案为 x_2。

算法2:假定各会诊专家的权重不等,即 $w=\{0.4,0.1,0.2,0.3\}$。

① 根据式(8-11)~式(8-14)将4种不同偏好信息一致化为非负和归一化的效用值。

$$U^1=\{0.3000,0.4000,0.1000,0.2000\}$$



$$U^2 = \{0.2000, 0.3000, 0.4500, 0.0500\}$$
$$U^3 = \{0.2167, 0.2750, 0.3083, 0.2000\}$$
$$U^4 = \{0.0546, 0.5688, 0.1151, 0.2615\}$$

② 将计算出的一致化效用值以及个各专家权重代入式(8-35)，计算出最优化模型的全局最优解 $\boldsymbol{U}_g^*$，即

$$u_1^* = 0.1793,\quad u_2^* = 0.4437,\quad u_3^* = 0.1681,\quad u_4^* = 0.2089$$

③ 根据计算出的最优群体效用值的大小对决策方案集 X 进行排序得如下结果：

$$x_2 \succ x_4 \succ x_1 \succ x_3$$

因此最优方案为 x_2。

由以上结果可以看出，两种方法计算的最满意方案是相同的。但是专家的权重对方案的排序产生了一些影响。

8.7　小结

本章对群决策的偏好表示、偏好集结函数和方法进行了分析和比较。结合鱼病远程会诊过程中会诊专家针对备选方案集给出的偏好信息的特点，选用了基于不同形式偏好信息的群决策相对熵集结最优化模型来求解。通过案例分析，表明该方法应用于鱼病远程会诊群决策是有效的。

参考文献

1. Chiclana F, Herrera F, Herrera-Viedma E. Integrating multiplicative preference relations in a multipurpose decision-making model based on fuzzy preference relations [J]. Fuzzy Sets and Systems, 2001, (122): 277～291
2. 徐泽水. 不确定多属性决策方法与应用[M]. 北京：清华大学出版社，2004
3. 樊治平，姜艳萍. 基于 OWG 算子的不同形式偏好信息的群决策方法[J]. 管理科学学报，2003，6(1)：32～36
4. 宋光兴，杨德礼. AHP 判断矩阵与模糊判断矩阵相互转化方法[J]. 大连理工大学学报，2003，43(4)：535～539
5. Herrera-Viedma E, Chiclana F, and Herrera F. A note on the reciprocity in the aggregation of fuzzy preference relations using OWA operators, Fuzzy Sets and Systems, 2003, 137 (1): 71～83
6. Delgado M, Herrera F et al. Combining Numerical and Linguistic Information in Group Decision Making [J]. Information Sciences, 1998, 107(2):177～194
7. 徐泽水. 模糊互补判断矩阵排序的一种算法[J]. 系统工程学报，2001，16(4)：311～314
8. 吴江. 群决策中 4 种偏好信息的转换方法研究[J]. 武汉理工大学学报，2004，26(3)：64～67
9. 王欣荣，樊治平. 一种具有不同形式偏好信息的群决策方法[J]. 东北大学学报，2003，24 (2)：178～181
10. 肖四汉，樊治平，王梦光. 群决策中两类偏好信息——AHP 判断矩阵和模糊偏好关系矩阵的一致化方法[J]. 系统工程学报，2002，17(1)：82～86
11. 樊治平，姜艳萍，肖四汉. 基于 OWA 算子的不同形式偏好信息的群决策方法[J]. 控制与决策，2001，16(增刊)：749～752
12. Chiclana F, Herrera F, Herrera-Viedma E. Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations[J]. Fuzzy Sets and Systems, 1998, 97(1):33～48

13. Herrera F, Herrera-Viedma E. Linguistic decision analysis: Steps for solving decision problems under linguistic information [J]. Fuzzy Set and Systems, 2000, 115(1):68～82
14. Herrera F, Herrera-Viedma E, Verdegay J L. Choice processes for non homogeneous group decision making in linguistic setting [J]. Fuzzy Sets and Systems, 1998, 94(3):287～308
15. Herrera F, Herrera Viedma E, et al. Multiperson decision making based on multiplicative preference relations [J],European Journal of Operational Research, 2001,129(2):372～385
16. 陈华友,刘春林.群决策中基于不同偏好信息的相对熵集成方法[J].东南大学学报(自然科学版),2005,35(2):311～315
17. 徐泽水.一种基于互反判断矩阵的多属性决策信息集成方法[J].系统工程,2002,20(2):93～96
18. 徐泽水,顾红芳.混合判断矩阵的两种排序方法[J].系统工程与电子技术,2002,24(5):1～3
19. 邱菀华.管理决策与应用熵学[M].北京:机械工业出版社,2002

第 9 章

基于证据理论的不确定性 FDT 群决策方法

在鱼病实际会诊过程中，由于时间限制和数据缺乏以及问题的非结构化，经常会出现会诊决策信息的不确定性或模糊性，从而给最终的一致性确诊带来了很大困难。专家经常能给出的决策信息是备选方案集和各方案的可信度。针对此问题，本章采用人工智能中处理不确定性信息的证据理论来对冲突意见进行信息融合，得出方案排序，提高群决策效果。

9.1 经典 D-S 证据理论的基本概念及性质

常规的决策分析理论以概率论和数理统计为基础，该理论认为概率是由事件发生的频率(作为证据)完全决定的，是纯客观的，该理论片面强调证据的作用，忽视人的判决作用。而 Bayesian 主观概率理论认为，概率是人的偏好或主观意愿的度量，是纯主观的，即该理论片面强调人的判决作用，而忽视客观证据的作用。证据理论认为，对于概率推断的理解，我们不仅要强调证据的客观化而且也要重视证据估计的主观性，概率是人在证据的基础上构造出的对一命题为真的信任程度，简称为信度。因此，证据理论可以根据各种资料对系统各个部分状态的概率进行归纳与估计，并作出正确的决策。证据理论是一种关于判决的理论，是一种不同于传统贝叶斯理论的新的概率解释方法，用以描述、表示并融合不确定信息。

最早对证据理论作出贡献的是 A. P. Dempster，他在 1967 年提出上下概率的概念，第一次给出了不满足可加性的概率。第二年，Dempster 针对统计问题给出了两批证据合成的原则，这成为后来证据理论中对两个独立信息源进行合成的基本法则。之后，G. Shafer 继续了 Dempster 的研究，并于 1976 年出版了《证据的数学理论》一书，标志着证据理论的诞生。因此，证据理论也被称为 Dempster-Shafer 理论。此后，除了 Shafer 本人继续做了很多工作以外，诸如 D. Dubois，H. Prade，P. Smits 和 Zadeh 等一批学者也参与到证据理论的研究中来，引入模糊数学的思想方法，进一步完善了证据理论。[1,2]

证据理论提出后，理论上有了逐步完善，而在实践中也取得了很多成功的应用。目前其应用领域主要在决策、预测、人工智能和专家系统等方面。从解决问题的角度来看，证据理论的一个重要优势就在于对多信息源不确定信息的表述和融合，以得到更为准确有效的信息。目前进行信息融合的数学工具主要有贝叶斯的概率理论、证据理论、模糊集论和神经网络等。H. Zhu 和 O. Basir 对如何运用证据理论进行信息融合提出了一般性的原则和方法，并通过一些仿真测试验证其有效性。而通过这种基于证据理论的信息融合办法，很多研究者又将其应用于各个领域，如遥感、医疗诊断、多准则决策以及不确定性推理等。[3,4]

国内从 20 世纪八九十年代开始也对证据理论进行理论和应用初探，积累了一定的研究成果。在理论方面的研究多在于对 Dempster 合成法则的探讨和改进以及将证据理论与其他数

学工具进行结合的尝试。证据理论在模式识别、故障诊断、多分类器融合、专家意见综合、决策以及风险评估等领域中都得到了较好的应用。

由于本章公式较多，故列出证据理论合成公式的主要公式关系框图，如图 9-1。

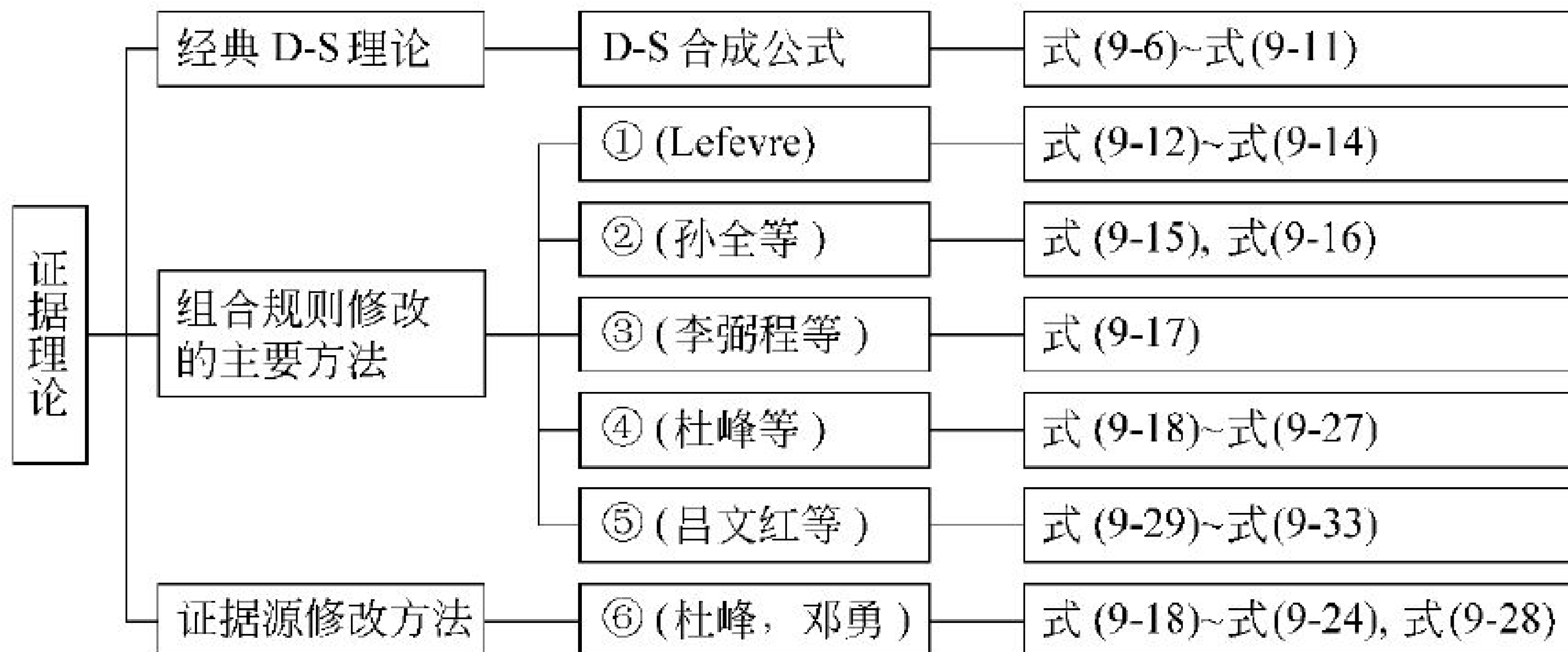


图 9-1　主要公式关系框图

9.1.1　概率的四种解释及性质

客观解释(频率解释)：概率描述了一个可以重复出现的事件的客观事实，即该事件可重复出现的频率。如掷色子。

个人主义解释(主观解释、贝叶斯解释)：概率反映了个人主义的一种偏好，是个人的主观意愿作用的结果。如赌博。

必要性解释(逻辑主义解释)：把概率看成命题与命题之间联系程度的度量。这种联系是纯客观的，与人的作用无关。演绎推理是概率推理的特例。

频率主义解释和逻辑主义解释赋予概率一种客观属性，概率的得到与人类主观活动没有关系。贝叶斯解释把概率解释成人的偏好或主观意愿的度量，而没有强调概率如何构造。[5]

构造性解释：Shafer 指出以上三种解释都没有涉及概率推断的构造性特征，证据理论给概率一种新的构造性解释，认为概率是某人在证据的基础上构造出的他对一命题为真的信任程度，简称信度。

前三种概率满足可加性：

$$\forall A, B \in \Theta,\quad A \cap B = \varnothing,\quad P(A \cup B) = P(A) + P(B) \tag{9-1}$$

根据可加性，如果我们相信一个命题为真的程度为 S，那我们就必须以 $1-S$ 的程度去相信该命题的反。在许多情况下，这是不合理的。例如对“地球以外存在着生命”这一命题。

Daldey 的不可能性定理指出：在以下两个基本假设下，①个体的概率估计 $P_i(E)$是独立随机变量；②不论给定的事件 E 发生或不发生，$P_i(E)$是独立随机变量；不存在一种集结个体概率估计的数学公式满足概率定理。即如果群体概率估计是个体概率估计的函数，则群体概率估计不满足概率定律，这意味着基于概率的集结模型在 Daldey 不可能性定理的条件下都将失效。

证据理论舍弃了这一原则，而用半可加性原则来代替。从而使 D-S 证据合成公式将多个证据合成后得到的不确定性度量依然满足证据基本可信数的性质，从而使证据的信度合成具有系统完整性。

9.1.2 基本定义

在证据理论中，证据指的是人们分析命题求其基本可信数所依据的事物的属性与客观环境，还包括人们的经验、知识和对该问题所作的观察和研究。人们通过对证据的分析得出命题的基本可信数分配函数 $m(A)$，如图9-2所示。

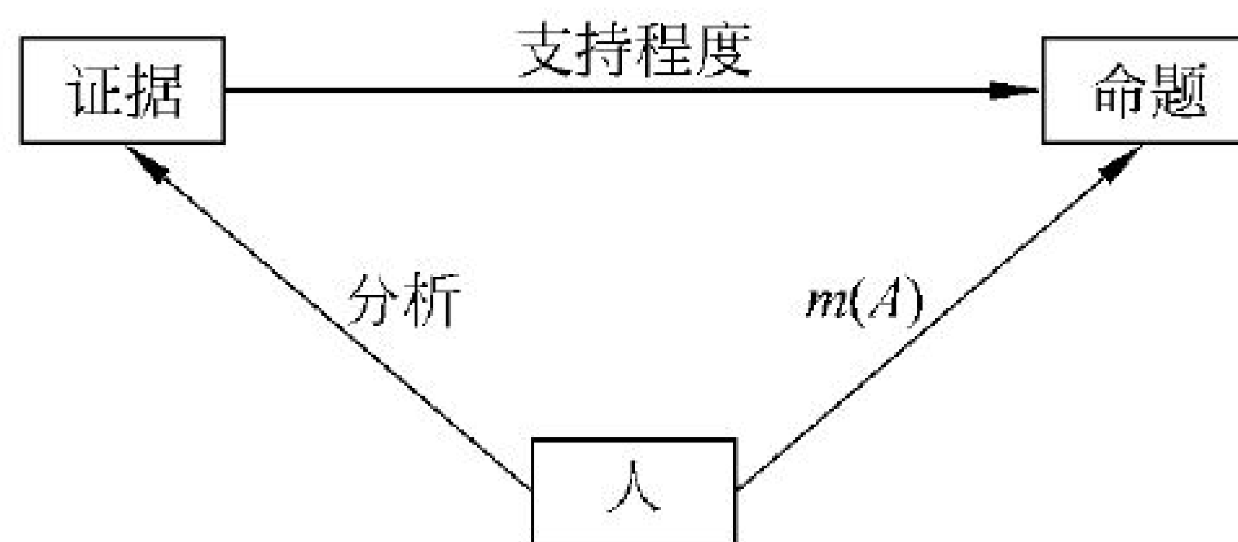


图9-2 信度的形成

对一个判决问题，设所能认识到的可能结果用集合 Θ 表示，那么人们所关心的任一命题都对应于 Θ 的一个子集。例如，对于鱼类肝脏疾病的诊断问题，可以根据以往的诊治记录和长期积累的经验知识，确定出如下的可能性集合：

Θ={急性肝细胞坏死型，亚急性肝细胞坏死型，慢性肝病型，胆小管型，潜伏型}

而弄清楚我们所关心的鱼到底得了什么样的肝病，也就是要在 Θ 中确定出一个子集。从以上这个例子可以看出，Θ 中每一个子集均对应于一个命题。将命题和子集对应起来，可以把比较抽象的逻辑概念转化为比较直观的集论概念。

Shafer指出，Θ 的选取依赖于我们的知识和认识水平。为了强调可能性集合 Θ 所具有的这种认识论的特性，Shafer把 Θ 称为识别框架(frame of discernment)。而且当一个命题对应于该识别框架的一个子集时，称为该框架能够识别的命题。另外，Θ 的选取应足够丰富，以便使所考虑的任何特定命题集都可以对应于 Θ 的幂集中的某个子集。若 $\Theta=\{\theta_1,\theta_2\}$，则有 Θ 的幂集：$2^\Theta=\{\varnothing,\{\theta_1\},\{\theta_2\},\Theta\}$。下面是一些基本概念的定义。[6]

定义9.1 设 Θ 为识别框架，如果集函数 $m:2^\Theta\to[0,1]$（2^Θ 为 Θ 的幂集）满足以下两个条件：

$$① \ m(\varnothing)=0; \quad ② \ \sum_{A\subseteq\Theta} m(A)=1 \tag{9-2}$$

则称 m 为框架 Θ 上的基本可信度分配(basic probability assignment)；$\forall A\subseteq\Theta$，$m(A)$ 称为 A 的基本可信数。基本可信数反映了对 A 本身的信度大小（而不去管它的任何真子集与前因后果）。条件①反映了对于空命题不产生任何信度；条件②说明了给所有命题赋予的信度数的和（总信度）等于1。

定义9.2 设 Θ 为识别框架，$m:2^\Theta\to[0,1]$ 为框架 Θ 上的基本可信度分配，则称由

$$\forall A\subseteq\Theta, \quad \mathrm{Bel}(A)=\sum_{B\subseteq A} m(B) \tag{9-3}$$

所定义的函数 $\mathrm{Bel}:2^\Theta\to[0,1]$ 为 Θ 上的信度函数(belief function)。$\mathrm{Bel}(A)$ 描述了对 A 的总信度。

定义9.3 $\forall A\subseteq\Theta$，若 $m(A)>0$，则称 A 为Bel的焦元。

定义9.4 设信度函数Bel的焦元为 $A_1,A_2,\cdots,A_k$，则称

$$C=A_1\cup A_2\cup\cdots\cup A_k \tag{9-4}$$

为Bel的内核。

定义9.5 设 $\mathrm{Bel}:2^\Theta\to[0,1]$ 为 Θ 上的信度函数，则称由

$$\forall A\subseteq\Theta, \quad \mathrm{pl}(A)=1-\mathrm{Bel}(\bar{A})=\sum_{B\cap A\neq\varnothing} m(B) \tag{9-5}$$

所定义的函数 $\mathrm{pl}:2^\Theta\to[0,1]$ 为Bel上的似真度函数。$\forall A\subseteq\Theta$，$\mathrm{pl}(A)$ 称为 A 的似真度。

pl(A)表示不怀疑 A 的程度或者说发现 A 可靠或似真的程度。

定义 9.6 $\forall A\subseteq\Theta$,称区间[Bel(A),pl(A)]为信度区间。

如前所述,Θ 中的每一个子集都对应与一个命题,故命题的不确定性可由集合的不确定性来表示,而 Bel(A)和 pl(A)分别给出了集合 A 的信度上限和下限值,因此,信度区间描述了命题的不确定性。图 9-3 直观地描述了命题的不确定性。

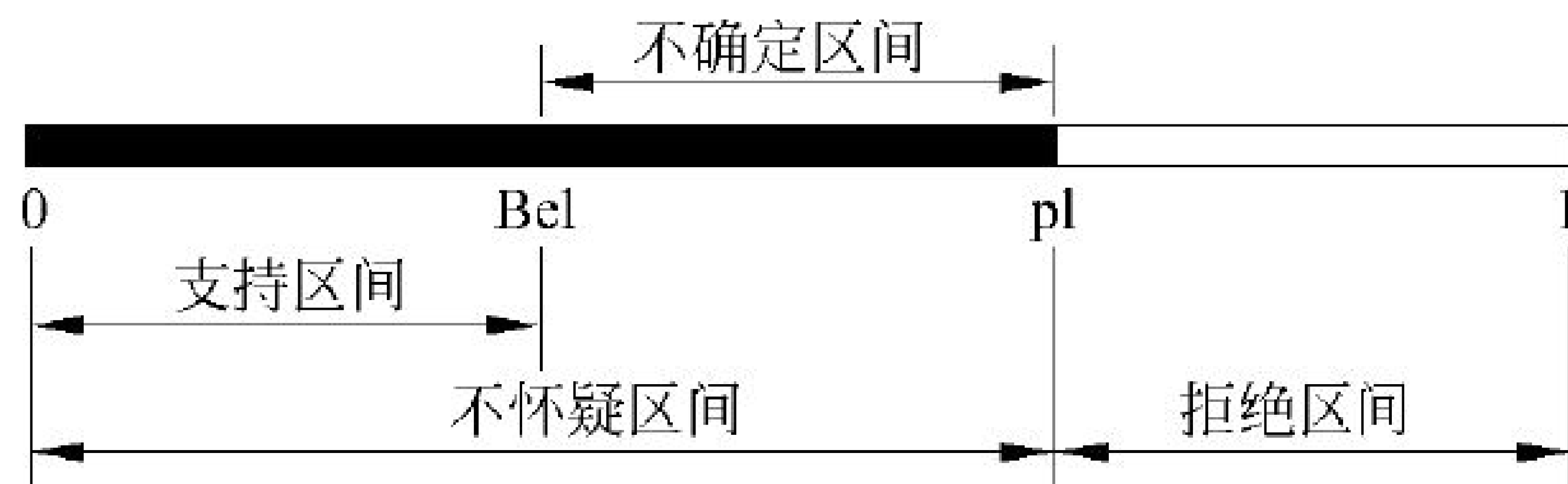


图 9-3　命题的不确定性表示

设 $a,b\in[0,1]$,令 Bel(A)$=a$,pl(A)$=b$,则信度区间$[a,b]$表示对 A 有一定程度的信任,也有一定程度的不信任;$[a,1]$表示对 A 有一定程度的信任,但无不信任;$[0,b]$表示对 A 有一定的不信任,但无信任;[1,1]表示 A 为真;[0,0]表示 A 为假;[0,1]表示对 A 完全不知道。

9.1.3　Dempster 合成规则及其性质

1. Dempster 合成规则

定理 9.1 设 Bel_1 和 Bel_2 是同一识别框架 Θ 上的两个信度函数,m_1 和 m_2 分别是其对应的基本可信度分配,焦元分别为 $A_1,A_2,\cdots,A_k,A_{k+1},\cdots,A_K$ 和 $B_1,B_2,\cdots,B_n,B_{n+1},\cdots,B_N$,设

$$\sum_{A_k\cap B_n=A} m_1(A_k)\cdot m_2(B_n)<1 \tag{9-6}$$

那么,合成后的基本可信度分配函数 $m:2^{\Theta}\to[0,1]$如下所示:

$$m(A)=\begin{cases}0, & A=\varnothing\\ \dfrac{\sum\limits_{A_k\cap B_n=A} m_1(A_k)\cdot m_2(B_n)}{\sum\limits_{A_k\cap B_n\neq\varnothing} m_1(A_k)\cdot m_2(B_n)}=\dfrac{\sum\limits_{A_k\cap B_n=A} m_1(A_k)\cdot m_2(B_n)}{1-k}, & A\neq\varnothing\end{cases} \tag{9-7}$$

其中:

$$k=\sum_{A_i\cap B_j=\varnothing} m_1(A_i)m_2(B_j) \tag{9-8}$$

由 m 所给定信度函数的核心等于 Bel_1 和 Bel_2 核心的交。由 m 给定的信度函数称为 Bel_1 和 Bel_2 的直和,记为 $\text{Bel}_1\oplus\text{Bel}_2$。如果式(9-6)不成立,那么就说 Bel_1 和 Bel_2 的直和不存在。

对两个证据合成来说,设 Bel_1 与 Bel_2 是识别框架 Θ 上的信度函数,用 Dempster 合成法则求它们的直和时,$A_i\cap B_j=\varnothing$时丢弃 $m_1(A_i)m_2(B_j)$。k 为在 Bel_1 与 Bel_2 合成的整个过程中所丢弃的总的信质。当 $A_i\cap B_j=\varnothing$时,Bel_1 与 Bel_2 分别给两个不相容的命题赋予了信度,两个证据发生了冲突。k 值越大,说明冲突越大。k 值反映了这两个证据间的冲突。

假设识别框架 $\Theta=\{C,D\}$,当证据 A 的基本可信数 m_a 为(0.4,0.6,0),当证据 B 的基本可信数 m_b 为(0.4,0.6,0),A, B 两个证据的合成为 m,由式(9-7)算得 m 为(0.3077,0.6923,0)。从该结果可看出当两个证据都对 D 事件具有较高的信度时,其合成结果就具有更高的信度。它较好地反映了当证据之间没有较大的冲突时人们的决策规律。



定理 9.2　设 $\mathrm{Bel}_1, \mathrm{Bel}_2, \cdots, \mathrm{Bel}_n$ 是同一识别框架 Θ 上的信度函数，$m_1, m_2, \cdots, m_n$ 是对应的基本可信度分配，如果 $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2 \oplus \cdots \oplus \mathrm{Bel}_n$ 存在，那么由下式定义的函数 $m: 2^{\Theta} \to [0,1]$ 称为合成后的基本可信度分配

$$m(A) = \begin{cases} 0, & A = \varnothing \\ \dfrac{\sum\limits_{\cap A_i = A} \prod\limits_{i=1}^{n} m_i(A_i)}{\sum\limits_{\cap A_i \neq \varnothing} \prod\limits_{i=1}^{n} m_i(A_i)} = \dfrac{\sum\limits_{\cap A_i = A} \prod\limits_{i=1}^{n} m_i(A_i)}{1-k}, & A \neq \varnothing \end{cases} \tag{9-9}$$

其中：

$$k = \sum_{\cap A_i = \varnothing} \prod_{i=1}^{n} m_i(A_i) \tag{9-10}$$

之所以进行证据体合成运算，是为了弥补单个证据体所具有的不确定性，提高证据推理结论的可靠性。

2. Dempster 合成规则的交换率和结合率

在 Dempster 合成规则中，$(n-1)$ 个证据合成可用下式表示：

$$m^{(n-1)}(A) = \frac{\sum\limits_{\cap A_i = A} \prod\limits_{i=1}^{n-1} m_i(A_i)}{\sum\limits_{\cap A_i \neq \varnothing} \prod\limits_{i=1}^{n-1} m_i(A_i)}$$

则结合率公式为

$$m^{(n)}(A) = \frac{\sum\limits_{\cap A_i = A} m^{(n-1)}(A) \cdot m_n(A_n)}{\sum\limits_{\cap A_i = \varnothing} m^{(n-1)}(A) \cdot m_n(A_n)} \tag{9-11}$$

由于集合的交运算具有交换律，容易证明 Dempster 合成法则具有交换律。

Dempster 合成方法的计算随着合成信息源 n 的增加，计算次数成指数增长。设 Θ 有 k 个元素，则 Dempster 合成规则的算法的复杂度为 $O(2^k \times k^n)$[7,8]。

Dempster 合成规则具有结合律可大大简化合成计算。可用递归调用两个证据的合成公式实现对 n 个证据的合成。并可监控计算过程，选择有利于证据合成质量提高的证据源。

9.2　证据理论合成规则的修正方法评述

Dempster 合成法则是一个反映证据联合作用的法则。给定几个同一识别框架 Θ 的基于不同证据的信度函数，如果这几批证据不是完全冲突的，那么就可以利用该合成法则计算出一个信度函数，而这个信度函数就可以作为那几批证据的联合作用下产生的信度函数。

尽管 D-S 证据理论可以有效地处理不确定性信息，但是在证据高度冲突情况下，经典 D-S 证据组合规则因归一化过程往往会产生有悖常理的融合结果。[9] 典型案例是两个医生看病的例子。对一个病人病情进行诊断，识别框架是 $\Theta = \{\theta_1, \theta_2, \theta_3\} = \{$脑膜炎，脑瘤，脑震荡$\}$，医生 A，B，分别给出如下信度值：

$$m_a(\theta_1)=0.99,\quad m_a(\theta_2)=0.01,\quad m_a(\theta_3)=0.00$$
$$m_b(\theta_1)=0.00,\quad m_b(\theta_2)=0.01,\quad m_b(\theta_3)=0.99$$

用 Dempster 合成规则合成，则可得脑瘤的信度值为 1。显然该合成结果不合常理，如何解决该问题成为完善证据理论的重要研究内容。鉴于此，很多学者在证据理论改进方面做了很多工作，提出了许多不同的改进思想和方法。通过分析，这些改进方法可以分为两类[18]，一类认为 Dempster 组合规则有问题，需要修改组合规则，如 Yager(1987)方法、Dubios(1998)方法等；另一类则认为组合规则没有问题，而给出的证据源需要修改，如 Murphy(2000)方法等。引起问题的关键原因是冲突信息的存在，但冲突信息的存在是很复杂的，上述两类思想实际上是从不同的角度处理冲突信息。在总结这两类修改思路的基础上，我们分析了相应的具有代表性的改进方法及其存在的问题，并对目前合成效果较好的方法进行了介绍。

9.2.1 修改组合规则

这类方法认为证据高度冲突下使用 Dempster 组合规则产生不合理结论是由该规则的归一化步骤所产生的。由 Dempster 组合规则中归一化过程可以看出，组合规则将冲突信息完全抛弃，在数学结果上就引出不合常理问题。为此，很多学者认为在组合规则中，冲突信息不能被简单地抛弃，而应重新分配，即组合规则需要修改。新的组合规则主要是解决如何重新分配冲突的问题，这一问题又可以分为：①冲突应该重新分配给哪些子集；②在确定可接收冲突的子集后，冲突应该以什么比例分配给这些子集。这一类解决方法的代表是 Lefevre 等人(2002)提出的修改方法。

Yager 将 Dempster 合成方法中归一化因子 $1/(1-k)$去掉，并将 k 赋给 $m(\Theta)$。这样做是将冲突的信度部分归入未知领域。Yager 公式对于不具冲突性的证据仍采用了与运算合成方式，但没有抛弃所有冲突性证据，而是把支持证据冲突的那部分概率全部赋予了未知领域。虽然 Yager 公式能合成高度冲突的证据，但由于对冲突性证据是完全否定的，因而在证据源多于两个时，合成效果有时并不理想。

Lefevre 等人提出的统一信度函数组合方法为

$$m(A)=\sum_{\substack{A_i,B_j\\A_i\cap B_j=A}} m_1(A_i)m_2(B_j)+m^c(A),\quad \forall A\subseteq\Theta \tag{9-12}$$

其中 $m^c(A)$表示冲突信息分配给子集 A 的集合，它又进一步写成：

$$\left.\begin{aligned} m^c(A)&=w(A,m)\cdot m(\varnothing), && \forall A\subseteq P\\ m^c(A)&=0, && \text{其他}\end{aligned}\right\} \tag{9-13}$$

其中，P 为冲突重新分配的子集的集合，而 $w(A,m)$是权重且

$$\sum_{A\subseteq 2^{\Theta}} w(A,m)=1 \tag{9-14}$$

它决定了分配给各个子集的冲突的大小。

通过设定冲突重新分配的子集的集合 P 和权重 $w(A,m)$，Lefevre 方法可以引申出该类解决思路中的其他方法，如 Demsper 组合规则，Smets 方法和 Yager 方法等。[11~13]

国内孙全等人认为即使证据之间存在着冲突，[14]它们也是部分可用的，并且可用程度取决于证据的可信度 ε。对 n 个证据，其合成公式如下：

$$m(A)=\begin{cases} p(A)+k\cdot\varepsilon\cdot q(A), & A\subset\Theta,A\neq\Theta\\ p(A)+k\cdot\varepsilon\cdot q(A)+k\cdot(1-\varepsilon), & A=\Theta\end{cases} \tag{9-15}$$



其中：

$$
\left.\begin{aligned}
p(A) &= \sum_{\substack{A_1,A_2,\cdots,A_n \\ A_1\cap A_2\cap\cdots\cap A_n=A}} m_1(A_1)m_2(A_2)\cdots m_n(A_n) \\
q(A) &= \frac{1}{n}\sum_{i=1}^{n} m_i(A) \\
\tilde{k} &= \frac{1}{n(n-1)/2}\sum_{1\leqslant i<j\leqslant n} k_{ij} \\
\varepsilon &= \mathrm{e}^{-\tilde{k}}
\end{aligned}\right\} \tag{9-16}
$$

$q(A)$为证据对 A 的平均支持程度；$\tilde{k}$ 反映了证据中两两冲突的平均程度；ε 反映了证据的可信度。这些表达都是较合理的。但有个问题是 ε 的定义有一定的主观因素。[13] 刘海燕等(2005)通过引入距离、相似度以及证据权重等概念对孙全的组合规则进行了改进。[15]

鉴于此，李弼程等提出一种更有效的组合规则，[16] 合成公式为

$$
\left.\begin{aligned}
m(\varnothing) &= 0 \\
m(A) &= p(A) + k \cdot q(A)
\end{aligned}\right\} \tag{9-17}
$$

其中 $p(A)$和 $q(A)$含义同式(9-16)，k 同式(9-8)。如此，证据冲突概率按各个命题的平均支持程度加权进行分配，具有明确合理的物理意义。[14]

杜峰，施文康和邓勇引入一个度量证据体间相似性程度的距离函数，基于冲突信息量可以部分或全部利用的思想提出了两个改进的组合规则。[17]

为了更合理地提取分歧意见之间的有效信息，吕文红等引入意见之间的距离、分歧、相似性测度等概念，针对群决策的特点，将冲突信息量的部分信任度归入未知领域，对合成规则进行了改进。[18]

9.2.2　证据源修改方法

Dempster 合成规则对各信息源提供的证据是平等对待的，认为各信息源提供证据的重要性与可靠性无优劣之分。而实际应用中往往会碰到不同的信息源提供证据的重要性与可靠性存在差异的情况，这时应用 Dempster 合成规则就需要修正。

这类解决方法的思路是：Dempster 组合规则本身没有错，在证据高度冲突时应该首先对冲突证据进行预处理，然后再使用 Dempster 规则。针对修改的组合规则，Haenni 指出，从实际应用角度看，都存在一个指定证据分配权重的问题，而且都不满足交换律和结合律，计算量问题将成为实际应用的瓶颈；从哲学角度出发，修改组合规则可看成“用 X 方法解决模型 Y，得到了违反直觉的结果 Z，认为方法 X 是错误的”。显然这是不科学的，因为有可能是模型 Y 的不适合。因此，部分学者认为，Dempster 的组合规则有其坚强的数学基础，是没有问题的，问题出在所给证据源上，需要通过修改降低冲突信息量后，再利用 Dempster 规则进行组合。典型的代表方法是 Murphy 的方法。

杜文吉基于集合加权“与”运算算子，给出加权的 Dempster 组合规则。该方法在一定的场合下比 Dempster 具有更好的性能，但是，该方法不具有互换律，合成时需按证据的优先级排序进行，而信息源优先级的确定依赖于该方法使用者的主观判断。梁昌勇等引入专家权威值来对基本可信数进行加权。杨春和李怀祖考虑不同信息、源与决策专家的重要程度，采用加权的方法在应用 Dempster 合成方法前对基本可信数 m 进行修正，然后使用 Dempster 合成方

法进行合成。

Murphy 的方法就是一种修改模型而不变 Demster 组合规则的方法。Murphy 分析了已有的改进方法，提出了一种证据平均组合规则。与其他方法相比较，该组合规则可以处理处突证据，且收敛速度较快。但是 Murphy 的平均方法只是将多源信息进行简单的平均，没有考虑各个证据之间的相互关联，这是该方法的不足之处。Jousselme 提出证据距离及两个证据源之间的相似度概念。其研究思想为：证据之间的距离在某种意义上反映证据的支持程度，两者距离越小，说明它们判断越接近。而且，两者之间相似的证据越多，说明证据的判断越接近事物的本质，证据源的整体可信度越高。

朱卫东指出传给 Θ 的基本可信数的大小取决于对基本可信数的综合修正系数。取值越小，传给未知领域 Θ 的基本可信数就越大。一般情况下，其取值面向具体应用问题，通过自适应学习、分析计算得出。例如，神经网络、遗传算法、模拟退火等计算方法都是寻找最优综合修正系数的有力工具。

基于这种思想，邓勇等(2004)在 Murphy 方法的基础上，引入一个度量证据体间相似性程度的距离函数，并进一步获得系统中各个证据被其他证据所支持的程度，将该支持度作为证据的权重，对多源证据进行加权平均后再利用 Dempster 组合规则融合证据信息。他提出的方法继承了 Murphy 方法的所有优点，并且具有更强的抗干扰能力，收敛速度更快。

这些方法都在一定程度上改进了冲突证据的合成效果。

9.2.3 各种改进方法的比较

从以上研究分析可以看出，杜峰等在总结这两类修改思路的基础上，分析了几种代表性的改进方法，分别从修改证据源和修改组合公式两种修改策略，以及修改组合公式时冲突信息量可以部分或全部利用的思想，提出了三种改进方案，并进一步分别进行了数值实验。理论分析和数值实验结果表明，改进后的三种方法不仅能充分体现相应的改进思想，而且可以更好地实现证据组合，较快地收敛真实目标，得到合理的融合结果。但不足之处是没有对两种修改策略进行数值仿真验算。吕文红等将证据理论引入群决策，并基于冲突信息量可以部分利用提出了修改证据组合规则的改进公式，但不足之处在于没有将该改进方法与现有方法作对比。

因此，下面对吕文红等和杜峰等的方法进行简单介绍与分析。

1. 证据理论改进方法一

杜峰，施文康和邓勇引入一个度量证据体间相似性程度的距离函数，提出了三种改进方案。

定义 9.7　Θ 为一包含 N 个两两不同命题的完备辨识框架，$E_{P(\Theta)}$ 是 Θ 所有子集生成的空间，$\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 是在识别框上的两个基本概率分配(BPA)，则 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 的距离可以表示为

$$d_{\mathrm{BPA}}(\boldsymbol{m}_1,\boldsymbol{m}_2)=\sqrt{\frac{1}{2}(\|\boldsymbol{m}_1(A_i)\|^2+\|\boldsymbol{m}_2(B_j)\|^2-2\langle\boldsymbol{m}_1(A_i),\boldsymbol{m}_2(B_j)\rangle)} \tag{9-18}$$

其中 $\|\boldsymbol{m}_i\|^2=\langle\boldsymbol{m}_i,\boldsymbol{m}_i\rangle$，$\langle\boldsymbol{m}_1,\boldsymbol{m}_2\rangle$ 为两个向量的内积。

$$\langle\boldsymbol{m}_1,\boldsymbol{m}_2\rangle=\sum_{i=1}^{2^N}\sum_{j=1}^{2^N}\boldsymbol{m}_1(A_i)\cdot\boldsymbol{m}_2(B_j)\frac{|A_i\cap B_j|}{|A_i\cup B_j|},\quad A_i,B_j\in P(\Theta) \tag{9-19}$$

设系统收集到证据为 n，可以利用上式计算出任两个证据体 $\boldsymbol{m}_i$ 和 $\boldsymbol{m}_j$ 之间的两两证据距离，得到距离矩阵：



$$\mathbf{DM} = \begin{bmatrix} 0 & \cdots & d_{1j} & \cdots & d_{1n} \\ \vdots & 0 & \vdots & & \vdots \\ d_{i1} & \cdots & 0 & \cdots & d_{in} \\ \vdots & & \vdots & 0 & \vdots \\ d_{n1} & \cdots & d_{nj} & \cdots & 0 \end{bmatrix} \tag{9-20}$$

由此定义证据体 $\boldsymbol{m}_i$ 和 $\boldsymbol{m}_j$ 之间的相似度

$$\mathrm{Sim}(\boldsymbol{m}_i, \boldsymbol{m}_j) = 1 - d_{\mathrm{BPA}}(\boldsymbol{m}_i, \boldsymbol{m}_j), \quad i, j = 1, 2, \cdots, n \tag{9-21}$$

其结果可表示为一个两个证据体之间的相似度矩阵：

$$\mathbf{SM} = \begin{bmatrix} 1 & \cdots & S_{1j} & \cdots & S_{1n} \\ \vdots & 1 & \vdots & & \vdots \\ S_{i1} & \cdots & 1 & \cdots & S_{in} \\ \vdots & & \vdots & 1 & \vdots \\ S_{n1} & \cdots & S_{nj} & \cdots & 1 \end{bmatrix} \tag{9-22}$$

在此指定系统中证据体 m_i 的支持度

$$\mathrm{Sup}(\boldsymbol{m}_i) = \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Sim}(\boldsymbol{m}_i, \boldsymbol{m}_j) \tag{9-23}$$

求出一个证据体 m_i 的支持度后，可以获得证据体 m_i 的可信度

$$\mathrm{Cred}_i = \frac{\mathrm{Sup}(\boldsymbol{m}_i)}{\sum_{i=1}^{n} \mathrm{Sup}(\boldsymbol{m}_i)}, \quad i, j = 1, 2, \cdots, n \tag{9-24}$$

可以看出 $\sum_{i=1}^{n} \mathrm{Cred}_i = 1$。

定义了一个群体可信度 ε 来表示现有证据源群体从整体上对理想证据群体的逼近程度，以作为划分冲突信息的依据。表示如下：

$$\varepsilon = \sqrt{\sum_{i=1, i<j}^{n} \mathrm{Sim}^2(\boldsymbol{m}_i, \boldsymbol{m}_j)} \bigg/ \sqrt{C_n^2 \times 1^2} \tag{9-25}$$

基于以上定义和前文所述的修改策略，提出了下面三种改进方案：

（1）冲突信息可以部分利用

$$\left.\begin{aligned} & m(A) = m_{\cap}(A) + m(\varnothing)\varepsilon \sum_{i=1}^{n} (m_i(A)\mathrm{Cred}_i), \quad \forall A \subseteq \Theta \\ & m(\Theta) = m(\Theta) + m(\varnothing)(1 - \varepsilon) \end{aligned}\right\} \tag{9-26}$$

（2）冲突信息全部可以分配

$$\left.\begin{aligned} & m(A) = m_{\cap}(A) + m(\varnothing) \sum_{i=1}^{n} (m_i(A)\mathrm{Cred}_i), \quad \forall A \subseteq \Theta \\ & m(\Theta) = 0 \end{aligned}\right\} \tag{9-27}$$

（3）修改证据源

$$\mathrm{mae}(m) = \sum_{i=1}^{n} (\mathrm{Cred}_i \times m_i) \tag{9-28}$$

将修改后的证据运用 Dempster 规则进行($n-1$)次组合。

杜峰等对以上三种改进方案分别进行了数值实验仿真验算，并与现有的其他算法进行了

比较。对比结果表明,这三种改进方案均优于同类算法,即冲突信息量部分可用的方法(式(9-26))优于孙全等提出的方法,冲突信息量全部可用的方法(式(9-27))优于李弼程等提出的方法。并且冲突信息量全部可用的方法(式(9-27))在融和结果上优于冲突信息部分可用式(9-26)的方法。但是在数值分析时没有对两种修改策略进行数值分析比较。我们在此基础上对两种策略(式(9-27)和式(9-28))进行了对比。

例 1 设辨识框架 $\Theta=\{A,B,C\}$,证据源:

$$S_1: m_1(A)=0.98,\quad m_1(B)=0.01,\quad m_1(C)=0.01$$
$$S_2: m_2(A)=0,\quad m_2(B)=0.01,\quad m_2(C)=0.99$$
$$S_3: m_3(A)=0.9,\quad m_3(B)=0,\quad m_3(C)=0.1$$
$$S_4: m_4(A)=0.9,\quad m_4(B)=0,\quad m_4(C)=0.1$$

可以看出,S_1,S_3,S_4 都在很大程度上支持 A,S_2 支持 C。

两种合成公式的结果对比如表 9-1 所示。此处用式(9-28)进行融合,为了降低计算量,在保证分辨率和合理性的情况下,各种情况只用 Dempster 规则融合了 1 次,并将融合结果与式(9-27)的计算结果进行了对比。从表中结果可以看出,修改证据源的方法的融合效果要优于修改组合规则的方法,而且减少了计算量。

表 9-1 融合结果对比 1

组合规则	证　据	$m(\varnothing)$	$m(A)$	$m(B)$	$m(C)$	$m(\Theta)$
式(9-27)	S_1,S_2	0.99	0.4851	0.01	0.5049	0
	S_1,S_2,S_3	0.999 01	0.881	0.0051	0.1139	0
	S_1,S_2,S_3,S_4	0.999 901	0.8902	0.0034	0.1064	0
式(9-28)	S_1,S_2		0.4898	0.0002	0.5100	
	S_1,S_2,S_3		0.9590	0.0001	0.0409	
	S_1,S_2,S_3,S_4		0.9859	0.0000	0.0141	

2. 证据理论改进方法二

吕文红等认为两意见之间的距离和这两个意见之间的分歧有着非常直接的关系,距离越大,分歧就越明显[15]。为了客观地衡量两意见之间的差异,定义两意见之间的距离如下:

定义 9.8 Θ 为一包含 N 个两两不同命题的完备辨识框架,E_{2^Θ} 是 Θ 所有子集生成的空间,m_1 和 m_2 是在识别框上的两个 BPA,则 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 的距离可以表示为

$$d_{\mathrm{BPA}}(\boldsymbol{m}_1,\boldsymbol{m}_2)=\frac{1}{2^N}\sqrt{\sum_{i=1}^{2^N}\sum_{j=1}^{2^N}\frac{\| A_i\cup B_j\|^2}{\| A_i\cap B_j\|^2}(m_1(A_i)-m_2(B_j))^2} \tag{9-29}$$

其中 $\|\cdot\|$ 为集合的基数。

在计算两意见距离时,各个幂集元素的大小的影响是不可忽视的,我们用其基数来衡量元素的大小。如果意见的 BPA 均是指派给单命题元素的,则 d 为一欧氏距离。

定义 9.9 Θ 为一包含 N 个两两不同命题的完备辨识框架,m_1 和 m_2 是在识别框上的任意两个 BPA,则 Θ 的距离的上确界为 $\mathrm{Sup}(d_{\mathrm{BPA}})$,$m_i$ 和 m_j 所表征的两意见之间的相似性测度为:

$$L(m_i,m_j)=2\mathrm{Sup}(d_{\mathrm{BPA}})-d_{\mathrm{BPA}}(m_i,m_j) \tag{9-30}$$

两意见的相似性测度是对两意见相似性的度量,但意见的相似性是一个相对的概念。在本方法中以距离上确界 $\mathrm{Sup}(d_{\mathrm{BPA}})$ 的两倍为参比值进行规整化处理,调节相似性测度的取值



范围,以适应不同的意见辨识框。

定义 9.10　Θ 为一包含 N 个两两不同命题的完备的辨识框架,$m_1, m_2, \cdots, m_n$ 是在识别框上的 n 个 BPA,m_i 所表征的意见的支持度为:

$$S(m_i) = \sum_{\substack{j=1 \\ i \neq j}}^{n} L(m_i, m_j) \tag{9-31}$$

定义 9.11　Θ 为一包含 N 个两两不同命题的完备的辨识框架,$m_1, m_2, \cdots, m_n$ 是在识别框上的 n 个 BPA,m_i 所表征的意见的相对信任度为:

$$P(m_i) = S(m_i) \Big/ \sum_{k=1}^{n} S(m_k) \tag{9-32}$$

由上述定义可以看出,$L(m_i, m_j)$反映的是两个意见的相似性程度,两个意见之间的距离越小,相似性测度 $L(m_i, m_j)$越大。$S(m_i)$反映的是某条意见被其他意见支持的程度。显然,当该意见与其他意见的相似测度的总和越大,就表明其他意见支持该意见的程度就越高,该意见就越可信。为了便于计算处理,对意见的支持度必须进行归一化处理,定义相对信任度 $P(m_i)$。

在群决策专家意见集结时,相对信任度大的意见对集结结果的影响应该较大。特别是在意见之间分歧较大时,应该分配给相对信任度较大的意见对集结结果影响较大的机会。同时,既然对分歧的意见无法作出合理的抉择,就应将其部分信任度归入未知邻域 X。由此可见,该方法也认为冲突信息量可以部分利用。由此,改进的证据组合公式应为:

$$\left.\begin{aligned} & m(\varnothing) = 0 \\ & m(A) = \sum_{A_i \cap B_j = A} m_1(A) m_2(B) + k \sum_{i=1}^{n} P_i m_i(A), \quad A \neq \varnothing, X \\ & m(X) = 1 - \sum_{i=1}^{n} m(A_i) \end{aligned}\right\} \tag{9-33}$$

其中,k 为两个冲突意见的冲突程度,计算同式(9-8)。

为了验证该方法的融合效果,我们进行了数值实验仿真验算,并与李弼程等提出的改进方法(参见式(9-17))进行了对比,如表 9-2 所示。

例 2　设辨识框架 $\Theta = \{A, B, C\}$,证据源:

$$\begin{aligned} & S_1: m_1(A) = 0.8, \quad m_1(B) = 0.2, \quad m_1(C) = 0 \\ & S_2: m_2(A) = 0, \quad m_2(B) = 0.2, \quad m_2(C) = 0.8 \\ & S_3: m_3(A) = 0.8, \quad m_3(B) = 0.1, \quad m_3(C) = 0.1 \\ & S_4: m_4(A) = 0.7, \quad m_4(B) = 0.1, \quad m_4(C) = 0.1 \end{aligned}$$

可以看出,S_1, S_3, S_4 都在很大程度上支持 A,S_2 支持 C。两种合成公式的结果如表 9-2 所示。

表 9-2　融合结果对比 2

组合规则	证　据	$m(\varnothing)$	$m(A)$	$m(B)$	$m(C)$	$m(\Theta)$
式(9-33)	S_1, S_2	0.96	0.128	0.104	0.128	0.64
	S_1, S_2, S_3	0.874	0.330	0.098	0.103	0.469
	S_1, S_2, S_3, S_4	0.739	0.551	0.083	0.115	0.251
式(9-17)	S_1, S_2	0.96	0.3840	0.2320	0.3840	0
	S_1, S_2, S_3	0.996	0.5312	0.1700	0.2988	0
	S_1, S_2, S_3, S_4	0.9996	0.5748	0.1503	0.2749	0

通过表 9-2 可知，吕文红等提出的方法效果不是很好。况且通过对以上计算过程的分析可知，该法在合成过程中使用了经典 D-S 合成规则的结合率，而实际上修改后的规则已经不满足结合率，所以此方法的正确性有待商榷。

综上所述，改进方法中的修改组合规则和修改证据源这两种修改策略，修改证据源策略从理论和实验分析来看都要优于修改组合规则。而且这种修改策略可利用 Dempster 合成规则的结合律，用递归调用两个证据的合成公式实现对 n 个证据的合成，大大简化合成计算。而且可以监控计算过程，如果融合结果已经可以正确分辨，就可以停止下一步的组合，而不一定必须组合($n-1$)次，提高了融合效率。

针对修改的组合规则，Haenni(2002)指出，从实际应用角度看，都存在一个指定证据分配权重的问题，而且都不满足交换律和结合律，随着证据体的增多，计算量问题将成为实际应用的“瓶颈”。对于只有两个证据合成的问题来看，计算是简便的。现有的这些方法中，杜峰等提出的冲突信息量可以全部利用的方法(式(9-27))的融合效果是最优的。但是该方法需要计算距离，这在证据很多时，计算量很大。而李弼程等提出的方法(式(9-17))虽然不及杜峰等提出的同类方法收敛速度快，但是计算相对简便，在证据体不多以及精度要求不高的情况下可以使用。

9.3 不确定性 FDT 的群决策流程

9.3.1 自然语言描述的基本可信度分配表示

多属性群决策是一类具有广泛实际背景的决策问题，它的一般性描述为：方案集 X 由有限个方案$\{x_1,x_2,\cdots,x_n\}(n\geqslant 2)$构成，有 k 个人参与决策，组成决策群体 $D=\{d_1,d_2,\cdots,d_k\}$，$k\geqslant 2$。群决策的目的就是集成所有成员的意见，对方案作出评价。

当决策参数(如方案效用)完全给定时，根据效用值的大小就可以确定方案的优劣。但是，人类对事物的认识，总是由浅入深的，对事物本质的认识，常常受到某种局限性的影响，特别是对发展中的事物或模糊事物的认识，更不会也不可能停留在一个固定的看法上。因此，在实际决策问题中，经常会出现决策信息具有不确定性或模糊性。另外由于时间的限制和数据的缺乏，问题的非结构化，也造成了人们在决策时表达偏好的不确定性。除此之外，在群决策情况下，由于各决策者的偏好不一致，也使得最终决策判断具有不确定性。例如，对于鱼类复杂病症来说，由于鱼病专家经验的有限性以及鱼病的复杂性，专家一般无法一次确诊。总是给出几个诊断假设以及各个假设的可信度。

由于在评估中有些指标难以定量化，因此可能采用“大致”，“可能”，“比较”等模糊词，这种表述正体现了专家对问题的准确理解，也更符合客观实际和人类的表达方式。目前多属性决策研究的一个热点就是如何使自然语言评价值的描述更能科学地表述不确定性信息。

我们将模糊词语集定义为：$U=\{$绝对肯定，很肯定，大致肯定，判断不出，大致否定，否定，绝对否定$\}$，用 7 个模糊词语表达不同的模糊程度，这 7 个模糊词语可取下值：$R=\{1.0,0.9,0.7,0.5,0.3,0.1,0\}$。实际上专家赋值时，并不拘于这 7 个数值，这些数值的作用只是希望引导专家的表达更加一致。

设 Θ 为一个辨识框 $A\subset\Theta$，有 k 个专家，第 i 个专家对集合 $A_{i1},A_{i2},\cdots,A_{ik}$ 的确认程度为 $r'_{i1},r'_{i2},\cdots,r'_{ik}$，令

$$r_{ij}=\frac{r'_{ij}}{\sum_{l=1}^{k}r'_{il}},\quad j=1,2,\cdots,k \tag{9-34}$$

则认为第 i 个专家所提出的评估意见在 Θ 上产生了一个基本可信度分配，$m_i: 2^{\Theta} \to [0,1]$，其中：

$$\left.\begin{aligned} m_i(A_{ij}) &= r_{ij}, \quad j = 1,2,\cdots,k \\ m_i(B) &= 0, \quad\quad B \neq A_{i1}, A_{i2}, \cdots, A_{ik} \end{aligned}\right\} \tag{9-35}$$

9.3.2 FDT群决策基本过程

基于D-S证据理论的不确定FDT群决策基本过程如图9-4所示，主要有以下步骤：

(1) 在鱼类复杂病症会诊群决策时，参与会诊的每个鱼病专家决策者在深入分析病症的基础上，可以凭自己的经验和知识或者借助鱼病诊断专家系统来构造系统的命题集(所有可能的几种诊断假设)，即系统识别框架 $\Theta=\{A_1, A_2, \cdots, A_k\}$。

(2) 针对待诊鱼病，构造基于识别框架的证据体 $E_i(i=1,2,\cdots,N)$，即各个专家的意见。

(3) 各专家针对所收集的资料以及自己的经验和知识，结合识别框架中各命题集合的特点，确定出各证据体(诊断假设)的基本可信度分配 $m_i(A_j)$ $(j=1,2,\cdots,K)$。

(4) 选用相应的合成公式对这些证据进行合成。

(5) 根据具体问题构造相应的决策规则，此处取“过半数原则”，即合成结果中存在可信度超过0.5的命题，而且合成结果与专家的直觉基本相符。如果所有命题的可信度都低于0.5，则需要会诊专家重新交流讨论，确定新的识别框架和证据体。

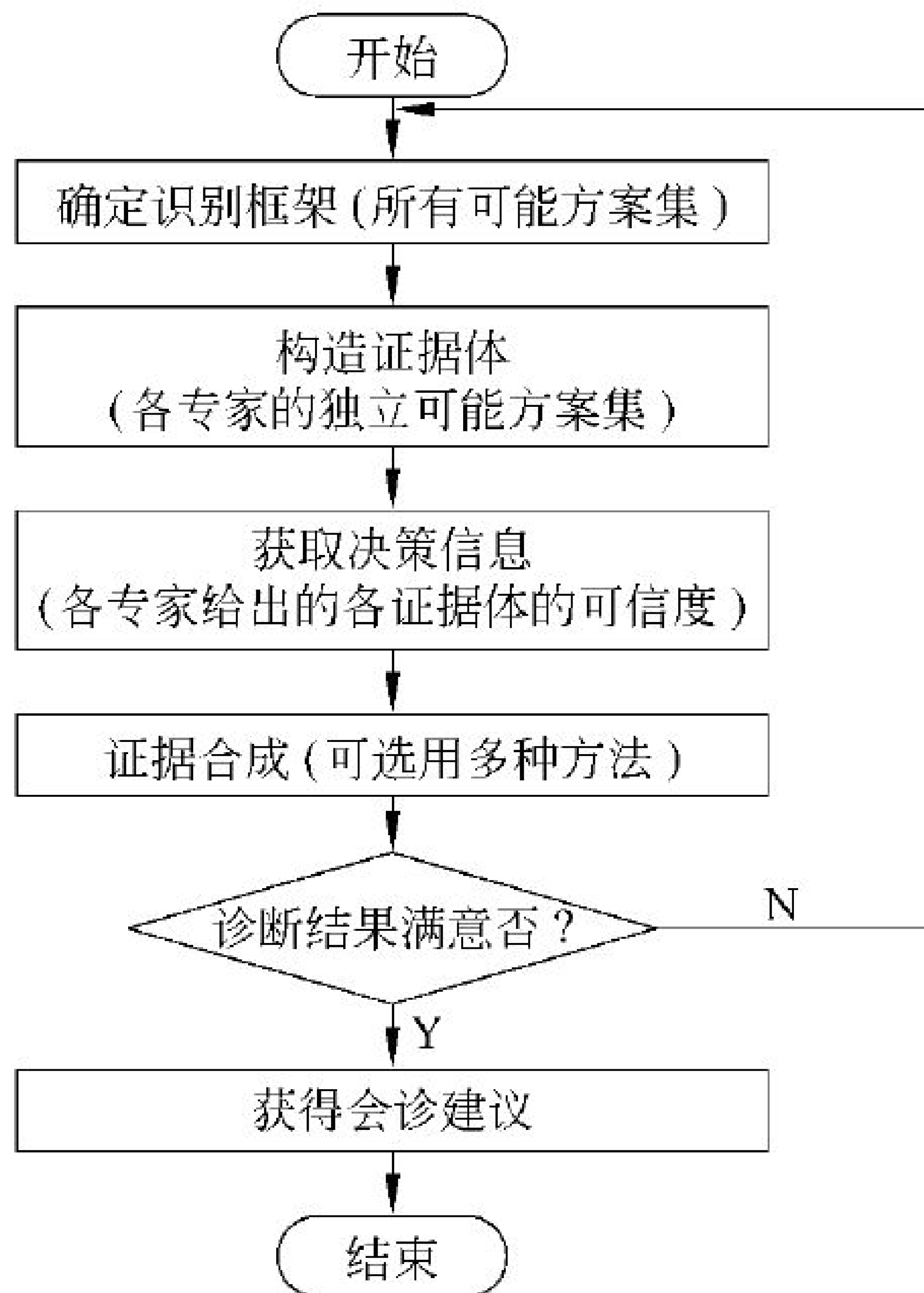


图9-4 基于证据理论的FDT群决策过程

(6) 根据该决策规则作出最终决策结论，进行治疗，在治疗过程中，不断依据病情以及新的信息修正治疗方案。

9.4 FDT实例分析

试析此例：遇到某个复杂病症会诊时，参与会诊决策的专家组可以凭借自己的经验和知识或者借助鱼病诊断专家系统提出诊断假设集，即本问题的识别框架{A(病毒性肠炎)，B(细菌性肠炎)，C(鲩内变形虫病)}，邀请4位专家给出各个诊断假设的可信度如下：

$$S_1: m_1(A) = 0.8, \quad m_1(B) = 0.2, \quad m_1(C) = 0$$
$$S_2: m_2(A) = 0, \quad m_2(B) = 0.2, \quad m_2(C) = 0.8$$
$$S_3: m_3(A) = 0.8, \quad m_3(B) = 0.1, \quad m_3(C) = 0.1$$
$$S_4: m_4(A) = 0.7, \quad m_4(B) = 0.1, \quad m_4(C) = 0.2$$

然后需要运用群决策方法给出一个会诊建议。

可以看出，S_1, S_3, S_4 都在很大程度上支持 A，S_2 支持 C。由于在群决策中，当各个专家的意见没有明显冲突时，可利用式(9-34)以及式(9-35)产生基本可信度分配，使用式(9-7)和

式(9-8)进行意见集结，会得到综合了所有专家意见的理想决策结果。但当两位或更多专家的意见存在明显分歧时，利用式(9-7)进行意见集结将会得到不合理的集结结果。在本例中，由式(9-7)得到将 S_1，S_2 两证据组合后的基本信任度分配函数如下：

$$m(A) = m(C) = 0$$

$$m(B) = 0.04/(1-(0.64+0.16+0.16)) = 1$$

由此可见，无论是专家1还是专家2都认为，得细菌性肠炎的可能性较小，而集结结果显示会诊决策结果就是细菌性肠炎，该会诊结果显然是不合理的。

因此根据杜峰和邓勇等(2004)的证据理论合成公式(见式(9-28))进行合成。

(1) 先根据式(9-18)、式(9-19)和式(9-20)计算距离矩阵

$$\mathbf{DM} = \begin{bmatrix} 0 & d_{12} & d_{13} & d_{14} \\ d_{21} & 0 & d_{23} & d_{24} \\ d_{31} & d_{32} & 0 & d_{34} \\ d_{41} & d_{42} & d_{43} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0.8 & 0.1 & 0.1732 \\ 0.8 & 0 & 0.7550 & 0.6557 \\ 0.1 & 0.7550 & 0 & 0.1 \\ 0.1732 & 0.6557 & 0.1 & 0 \end{bmatrix}$$

(2) 根据式(9-21)和式(9-22)计算相似度矩阵

$$\mathbf{SM} = \begin{bmatrix} 1 & s_{12} & s_{13} & s_{14} \\ s_{21} & 1 & s_{23} & s_{24} \\ s_{31} & s_{32} & 1 & s_{34} \\ s_{41} & s_{42} & s_{43} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.2 & 0.9 & 0.8268 \\ 0.2 & 1 & 0.2450 & 0.3443 \\ 0.9 & 0.2450 & 1 & 0.9 \\ 0.8268 & 0.3443 & 0.9 & 1 \end{bmatrix}$$

(3) 对前两个意见 S_1，S_2 进行集结。根据式(9-23)和式(9-24)计算各意见的可信度如下：

$$\mathrm{Cred}_1 = 0.5, \quad \mathrm{Cred}_2 = 0.5$$

根据式(9-28)计算加权平均后的意见为

$$m(A) = 0.4, \quad m(B) = 0.2, \quad m(C) = 0.4$$

由组合公式(9-7)和式(9-8)组合1次得到：

$$m(A) = 0.4444, \quad m(B) = 0.1111, \quad m(C) = 0.4444$$

(4) 对前三个意见 S_1，S_2，S_3 进行集结。根据式(9-23)和式(9-24)计算各意见的可信度如下：

$$\mathrm{Cred}_1 = 0.4089, \quad \mathrm{Cred}_2 = 0.1654, \quad \mathrm{Cred}_3 = 0.4257$$

根据式(9-28)计算加权平均后的意见为

$$m(A) = 0.6677, \quad m(B) = 0.1574, \quad m(C) = 0.1749$$

由组合公式(9-7)和式(9-8)组合1次得到：

$$m(A) = 0.8895, \quad m(B) = 0.0494, \quad m(C) = 0.0610$$

(5) 对前四个意见 S_1，S_2，S_3，S_4 进行集结。根据式(9-23)和式(9-24)计算各意见的可信度如下：

$$\mathrm{Cred}_1 = 0.2820, \quad \mathrm{Cred}_2 = 0.1155, \quad \mathrm{Cred}_3 = 0.2993, \quad \mathrm{Cred}_4 = 0.3031$$

根据式(9-28)计算加权平均后的意见为

$$m(A) = 0.6772, \quad m(B) = 0.1397, \quad m(C) = 0.1830$$

由组合公式(9-7)和式(9-8)组合1次得到

$$m(A) = 0.8961, \quad m(B) = 0.0381, \quad m(C) = 0.0654$$

从以上结果可以看出，最终会诊建议为 A(病毒性肠炎)，与直觉相符，所以合成结果是合



理的。因此用证据理论来集结会诊专家的意见是有效的。

9.5　小结

本章对证据理论进行了简要介绍，详细分析讨论了冲突证据的合成规则，对目前所用的改进方法进行了分析比较。将证据理论引入鱼病远程会诊群决策，通过案例分析表明，用证据理论以及冲突证据合成公式来处理鱼病会诊群决策问题是可行和有效的。

参考文献

1. Dempster AP, Laird NM, Rubin DB. Journal of the Royal Statistical Society. Series B (Methodological), 1977, 39(1): 1～38.
2. 牟克典，林作铨. 定性 Dempster-Shafer 理论[J]. 计算机研究与发展，2005，42(11)
3. Zhu H and Basir O. A scheme for constructing evidence structures in Dempster-Shafer Evidence Theory for data fusion[J]. IEEE International Symposium on Computational Intelligence in Robotics and Automation, 2003, Vol. 2: 960～965
4. Beynon M, Curry B, Morgan P. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modeling[J]. Omega, The International Journal of Management Science, 2000, (28): 37～50
5. Shenoy PP. and Shafer G, Propagating belief functions using local computations[J]. IEEE. Expert, 1986, 3(1), 43～52
6. 段新生. 证据理论与决策、人工智能[M]. 北京：中国人民大学出版社，1993
7. 朱卫东，杨善林，任明仑. 基于学习与证据理论的专家群体预测系统研究. 预测，2003，22(1)
8. 朱卫东. 面向互联网基于证据理论的智能决策支持系统研究[D]. 合肥工业大学，2003
9. Zadeh, Review of books: a mathematical theory of evidence[M]. 1987
10. Hegarat-Mascle SL, Bloch L and Vidal-Madjar D. Application of Dempster-Shafer Evidence Theory to unsupervised classification in multisource remote sensing[J]. IEEE Transactions on Geoscience and Remote Sensing, 1997, 35: 1018～1031
11. Smets P. The combination of evidence in the transferable belief model[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1990, 12(5): 447～458
12. Dubois D, Prade H. Representation and combination of uncertainty with belief functions and possibility measures [J]. Computational Intelligence, 1998, (4): 244～264
13. Ganna M, Rombaut M and Goutte R, et al. Improvement of brain lesions detection using information fusion approach[J]. IEEE International Conference on Signal Processing, 2002, (2): 1004～1007
14. 孙全，叶秀清，顾伟康. 一种新的基于证据理论的合成公式[J]. 电子学报，2000，(8): 117～119
15. 杜文吉，谢维信. 基于证据理论的模式识别算法[J]. 西安电子科技大学学报，1999，(8): 533～536
16. 李弼程，王波，魏俊等. 一种有效的证据理论合成公式[J]. 数据采集与处理，2002，(17): 33～36
17. 杜峰，施文康，邓勇. 证据特征提取及其在证据理论改进中的应用[J]. 上海交通大学学报，2004，38(S1): 164～168
18. 吕文红，吴祈宗，郭银景. 基于 D-S 证据理论的群决策专家意见集结方法[J]. 运筹与管理，2005，14(2): 10～13，25
19. 孙怀江，胡钟山，杨静宇. 基于证据理论的多分类器融合方法研究[J]. 计算机学报，2001，(3): 231～235
20. 柳毅，陈红林. 证据理论与模糊神经网络相结合的身份估计方法[J]. 火力与指挥控制，2003，28(1):

29～32
21. 蔡兴国，马平. 基于证据理论的并发故障诊断方法[J]. 哈尔滨工业大学学报，2003，35(4)：447～450
22. 鞠彦兵，冯允成，姚李刚. 基于证据理论的软件开发风险评估方法[J]. 系统工程理论方法应用，2003，12(3)：218～223
23. 梁昌勇，陈增明，黄永青等. Dempster-Shafer 合成法则悖论的一种消除方法[J]. 系统工程理论与实践，2005，(3)：7～12，85
24. 邓勇，施文康，朱振福. 一种有效处理冲突证据的组合方法[J]. 红外与毫米波学报，2004，23(1)：27～32
25. 卓越，王红，吴秋峰等. 群决策支持系统中的一致性分析技术[J]. 控制与决策，1999，14(6)：636～641
26. 张金玉，张优云，谢友柏. 基于证据理论的综合诊断理论及其应用[J]. 机械科学与技术，2000，(3)：183～186，206

第10章

基于粗糙神经网络的知识获取模型

本章通过对粗糙集(RS)理论和神经网络(ANN)知识获取方法的分析,结合鱼病案例知识的特点,建立了基于粗糙集和神经网络(RNN)松耦合的鱼病知识获取模型。

10.1 粗糙神经网络松耦合知识获取模型

10.1.1 模型基本思想

粗糙集(rough set,RS)与人工神经网络(artificial neural network,ANN)松耦合混合知识获取系统的基本思想是:将来自呼叫中心以及鱼病专家的原始鱼病诊断案例作为学习的样本集;将作为条件属性的鱼病症状数量化,并用(0,1)表示,按鱼病类别将大量鱼病案例分块建立决策表;利用粗糙集约简理论对条件属性进行相对约简,通过去掉冗余条件属性得到决策表的最小条件属性集。[1]

本章采用BP神经网络模型,以最小条件属性集对应的学习样本集作为神经网络的测试样本集,该样本集除去了所有不必要的条件属性,仅保留了影响鱼病分类的重要属性。用约简得到的最小条件属性集对应BP神经网络的输入和输出神经元进行学习和测试,输出获取结果。

10.1.2 模型构造过程

根据粗糙集理论和神经网络知识获取方法的要求,将鱼病诊断案例特征属性值数量化,然后组织建立鱼病诊断决策表,建立基于属性 $C \cup D$(C:鱼病症状集;D:鱼病集)的矩阵,计算出 $posC(D)$ 值以便对决策表进行属性约简,确定BP(back propagation)算法的神经元网络层数,输入信号阈值和权值。[2]鱼病粗糙神经松耦合知识获取模型如图10-1所示。

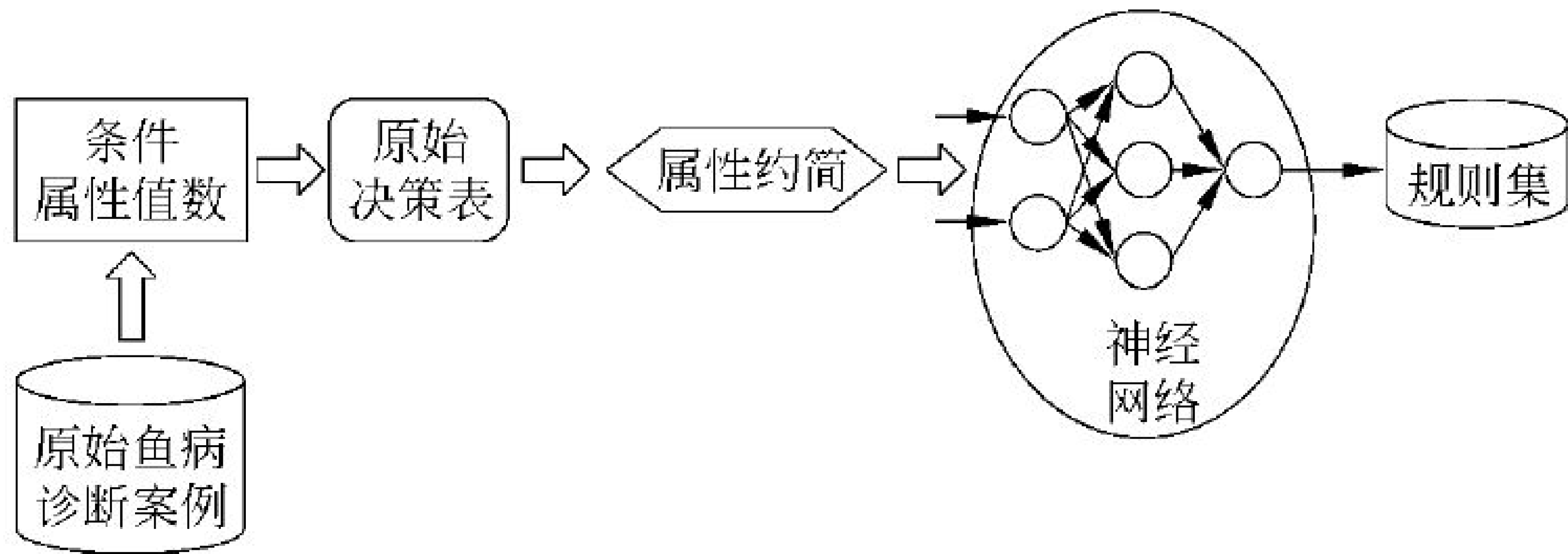


图10-1　鱼病粗糙神经松耦合知识获取模型

10.2 鱼病知识获取步骤

鱼病的 RS 与 ANN 松耦合混合系统：主要是粗糙集预处理神经网络系统，即用 RS 理论对 ANN 输入端的鱼病案例属性进行约简，寻找属性间关系，简化 ANN 结构。[3]该模块由两部分构成：粗集预处理和学习程序，粗集预处理是运用粗糙集理论对输入神经网络的数据进行前置处理，然后是采用 ANN 进行样本学习。整个过程包括以下四个步骤：

(1) 原始鱼病案例获取：获取原始鱼病诊断案例，以向量形式保存数据，行表示样本，列表示变量或属性；本研究中的鱼病原始案例大部分来自呼叫中心，这些案例都是用自然语言描述的，首先要将用自然语言表示的特征向量转化为相应的数据形式并以矩阵形式保存。

(2) 用决策表表示鱼病案例：原始鱼病诊断案例用决策表的形式表示，鱼病症状表示为条件属性值(C)，鱼病表示为决策属性值(D)，所有鱼病案例表示为对象(U)，V 表示属性的值域，则原始鱼病诊断案例表示为决策表的形式：$T=(U,C,D,V)$；采用二元信息表(0,1)，其中“0”表示没有此症状属性，“1”表示存在此种症状属性。

(3) 案例决策表约简：这一步要去掉鱼病案例决策表中的冗余特征属性而不改变决策表的分类效果。粗糙集理论提供了很多属性约简的方法，本研究主要采用通过计算条件属性(鱼病症状)的决策属性(鱼病种类)的正域来实现约简的方法。

(4) 神经网络的规则获取：将属性简约后得到的鱼病症状属性和决策属性作为神经网络的输入变量和输出变量，使神经网络结构得到简化，对样本进行训练和学习，最后得出鱼病诊断规则。

10.2.1 鱼病知识约简

对鱼病知识进行约简是知识获取的核心内容。显而易见，鱼病知识库中的所有知识(属性)并不是同等重要的，甚至其中某些知识是冗余的。而所谓知识约简，就是在保持知识库分类能力不变的条件下，删除其中不相关或不重要的知识。

鱼病知识约简涉及两个基本概念：约简和核。[4]讨论这两个概念前，先介绍几个定义。

定义 10.1 设 U 是一个论域，P 是定义在 U 上的一族等价关系，$R\in P$。如果 $\mathrm{ind}(P)=\mathrm{ind}(P-\{R\})$，则称关系 R 在 P 中是绝对不必要的(多余的)；否则，称 R 为 P 中必要的。绝对不必要的关系在知识库中是多余的，如果将它们从知识库中去掉，不会改变该知识库的分类能力。相反，若知识库中去掉一个绝对必要的关系，则一定会改变知识库的分类能力。

其中 $\mathrm{ind}(P)=\{(x,y)\in U\times U \mid \forall a\in P, f(x,a)=f(y,a)\}$ 称为属性集 P 的不可区分关系，如果 $(x,y)\in\mathrm{ind}(P)$，则称 x 和 y 是 P 不可区分的。容易证明 $\forall P\subseteq A$，不可区分关系 $\mathrm{ind}(P)$ 是 U 上的等价关系，$U/\mathrm{ind}(P)$(简记为 U/P)表示不可区分关系 $\mathrm{ind}(P)$ 在 U 上导出的划分。

定义 10.2 设 U 是一个论域，P 是定义在 U 上的一族等价关系，$R\in P$。如果每个关系 $R\in P$ 在 P 中都是绝对必要的，则称关系族 P 是独立的；否则，称 P 是相互依赖的。对于相互依赖的关系族来说，其中包含冗余关系，可以对其约简；而对于独立的关系族，去掉其中任何一个关系都将破坏知识库的分类能力。

定义 10.3 设 U 是一个论域，P 是定义在 U 上的一个等价关系族。P 中所有绝对必要关系组成的集合称为关系族 P 的绝对核，记作 $\mathrm{core}(P)$。



定义 10.4　设 U 是一个论域，P 和 Q 为定义在 U 上的两个等价关系族，且 $Q\subseteq P$。如果 $\mathrm{ind}(Q)=\mathrm{ind}(P)$，$Q$ 是独立的，则称 Q 是 P 的一个绝对约简，记作 $\mathrm{red}(P)$。如果知识 Q 是知识 P 的绝对约简，那么 U 中通过知识 P 可区分的对象，同样可以用知识 Q 来区分。

讨论鱼病决策表知识表达系统约简时，一个条件属性（鱼病症状）S 对应着一个等价关系（也称不分明关系或不可分辨关系），即在条件属性（鱼病症状）S 上取值的相等关系，它对论域 U 形成一个划分 U/S。决策表的所有条件属性形成条件属性集合 C 对论域的划分为 U/C，同时，决策属性集 $D=\{d\}$ 也对论域形成一个划分 U/D。

这两个划分形成了条件属性和决策属性在对论域样本 U 分类上的知识。鱼病属性约简的目标就是要从鱼病症状集合中发现部分必要的条件属性，使得根据这部分条件属性形成的相对于决策属性的分类和所有条件属性所形成的相对于决策属性的分类一致，即和所有条件属性相对于决策属性 D 有相同的分类能力。

定义 10.5　设 U 是一个论域，对于每个子集 $X\subset U$ 和一个等价关系 R，X 的正域 R 记为 $\mathrm{pos}_R(X)$，定义为：$\mathrm{pos}_R(X)=\bigcup\{r\in U/R: r\subseteq X\}$。

即正域表示根据知识（属性）R 判断肯定属于 X 的 U 中元素的集合。当 $r\in R$，且 $\mathrm{pos}_R(X)=\mathrm{pos}_{(R-\{r\})}(X)$ 时，称 r 为 R 中 X 不必要的，否则 r 为 R 中 X 必要的。

10.2.2　约简步骤和算法

鱼病决策表约简算法如下：输入决策表 DT(decision table)具有条件属性集（鱼病的症状集合）$C=\{C_1,C_2,\cdots,C_n\}$ 和决策属性集（鱼病集合）$D=\{D_1,D_2,\cdots,D_n\}$。

步骤 1：计算 D 的 C 正域 $\mathrm{pos}_C(D)$；

步骤 2：从 C 中剔除属性 C_i，$C_t=C-\{C_i\}$；

步骤 3：计算 D 的 C_t 正域 $\mathrm{pos}_{C_t}(D)$；

步骤 4：若 $\mathrm{pos}_{C_t}(D)=\mathrm{pos}_C(D)$ 则去除 C_i，反之保留 C_i；

步骤 5：按以上步骤遍历所有属性；

输出：条件属性集 C 对决策属性 D 的一个相对约简。

该算法的过程是：从鱼病症状属性集 C 中一个个地去除属性，如果去除该症状属性会造成对决策属性正域的变化，则保留该属性；否则去除该属性。最后剩下的鱼病症状属性集就是条件属性的一个约简。鱼病案例知识约简流程如图 10-2 所示。

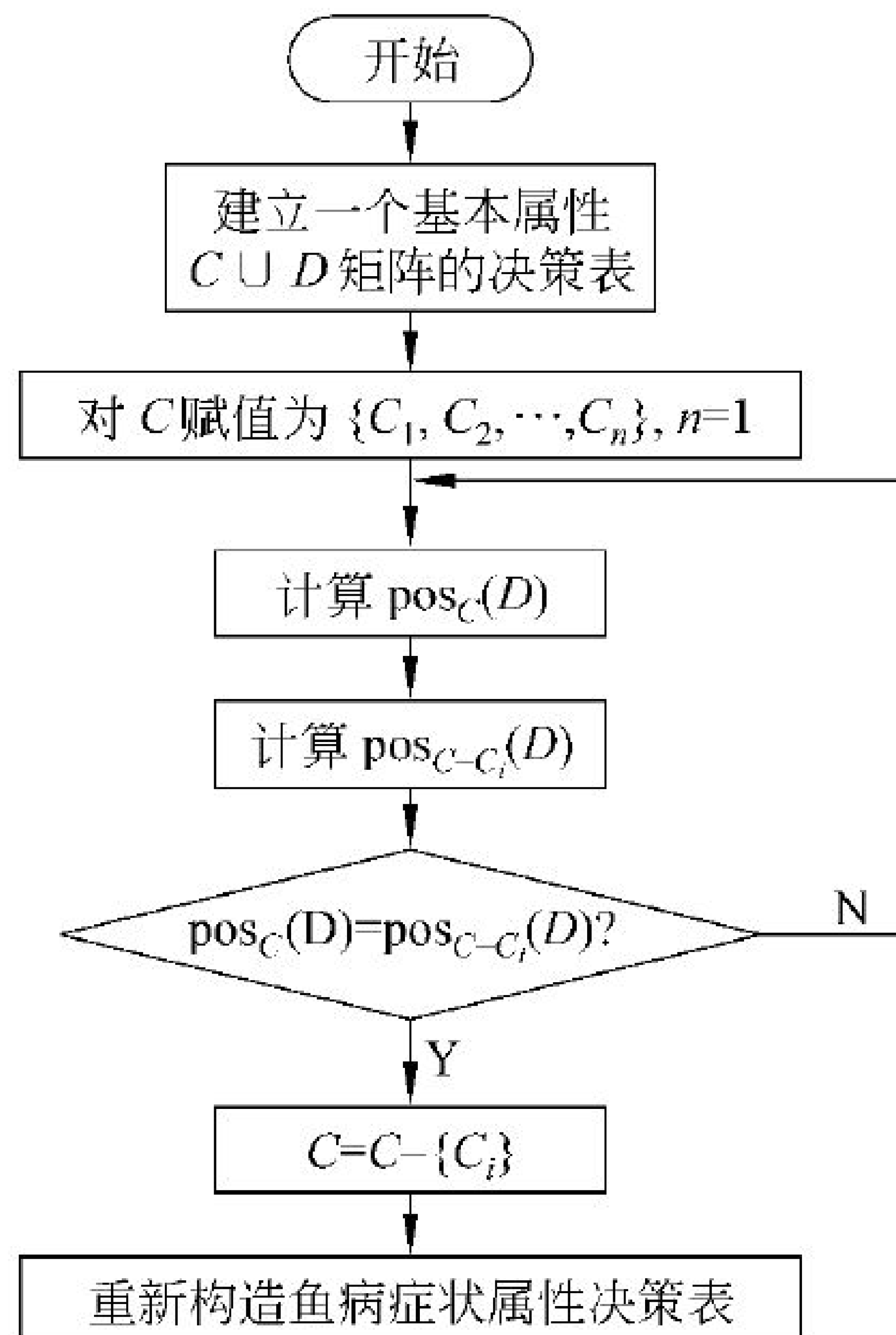


图 10-2　鱼病案例知识约简流程图

10.2.3　约简实例

以寄生虫鱼病诊断案例集为例，说明决策表约简的过程。表 10-1 所示为原始寄生虫鱼病案例集，其中 U 是案例集的有限集合；$A=C\cup D$，C 为条件属性（鱼病症状）的集合，其中 C_1 表示吃食不好；C_2 表示病鱼逛游；C_3 为头部发黑；C_4 为腮部肿胀；C_5 代表腮部充血。条件属性的值域为{0=没有这种症状，1=有这种症状}。

表 10-1　原始寄生虫鱼病案例集

U	C_1	C_2	C_3	C_4	C_5	D
x_1	0	1	1	0	0	0
x_2	0	1	0	0	1	0
x_3	1	1	0	1	0	1
x_4	1	1	1	0	0	1
x_5	1	0	0	1	1	0
x_6	1	1	0	1	1	1
x_7	0	1	1	0	1	1
x_8	0	0	1	0	1	1
x_9	1	1	0	1	1	1

由表 10-1 可以看出对象 x_6 和 x_9 属性的值域相同,即两个实例相同,因此根据属性约简的步骤,我们先删除决策表中重复的行,保留其一,得到表 10-2。

表 10-2　删除行的寄生虫鱼病案例集

U	C_1	C_2	C_3	C_4	C_5	D
x_1	0	1	1	0	0	0
x_2	0	1	0	0	1	0
x_3	1	1	0	1	0	1
x_4	1	1	1	0	0	1
x_5	1	0	0	1	1	0
x_6	0	1	1	0	1	1
x_7	0	0	1	0	1	1
x_8	1	1	0	1	1	1

在初步整理后的决策表中,对各个等价关系相对于论域 U 分类:

$$U/C_1 = \{\{x_1, x_2, x_6, x_7\}, \{x_3, x_4, x_5, x_8\}\}$$
$$U/C_2 = \{\{x_1, x_2, x_3, x_4, x_6, x_8\}, \{x_5, x_7\}\}$$
$$U/C_3 = \{\{x_1, x_4, x_6, x_7\}, \{x_2, x_3, x_5, x_8\}\}$$
$$U/C_4 = \{\{x_1, x_2, x_4, x_6, x_7\}, \{x_3, x_5, x_8\}\}$$
$$U/C_5 = \{\{x_1, x_3, x_4\}, \{x_2, x_5, x_6, x_7, x_8\}\}$$
$$U/\{C_1, C_2, C_3, C_4\} = \{\{x_1, x_6\}, \{x_3, x_8\}, x_2, x_4, x_5, x_7\}$$
$$U/\{C_1, C_3, C_4, C_5\} = \{x_1, x_2, x_3, x_4, \{x_5, x_8\}, \{x_6, x_7\}\}$$
$$U/\{C_2, C_3, C_4, C_5\} = \{\{x_1, x_4\}, x_3, x_2, x_5, x_8, x_6, x_7\}$$
$$U/\{C_1, C_2, C_3, C_5\} = \{x_1, x_4, x_3, x_2, x_5, x_8, x_6, x_7\}$$
$$U/\{C_1, C_2, C_4, C_5\} = \{x_1, \{x_2, x_6\}, x_3, x_4, x_5, x_7, x_8\}$$

计算分类 D 相对于分类 C 的依赖度:

$$U/C = \{x_1, x_4, x_3, x_2, x_5, x_8, x_6, x_7\}$$
$$U/D = \{\{x_1, x_2, x_5\}, \{x_3, x_4, x_8, x_6, x_7\}\}$$
$$\mathrm{pos}_C(D) = \{x_1, x_4, x_3, x_2, x_5, x_8, x_6, x_7\}$$
$$k = |\mathrm{pos}_C(D)| / |U| = 8/8 = 1$$

结果说明决策属性完全依赖于条件属性。



计算 $\text{pos}_C(D)$、$\text{pos}_{C-C_i}(D)$两者的关系，考察 C 中的条件属性是否可以省略，进而删除决策表中的列，从而求出条件属性的约简。

$$\text{pos}_{(C-C_1)}(D)=\{x_6,x_3,x_2,x_5,x_8,x_7\}\neq \text{pos}_C(D)$$

$$\text{pos}_{(C-C_2)}(D)=\{x_4,x_3,x_2,x_1,x_6,x_7\}\neq \text{pos}_C(D)$$

$$\text{pos}_{(C-C_3)}(D)=\{x_1,x_3,x_4,x_5,x_6,x_8,x_7\}\neq \text{pos}_C(D)$$

$$\text{pos}_{(C-C_4)}(D)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_8,x_7\}=\text{pos}_C(D)$$

$$\text{pos}_{(C-C_5)}(D)=\{x_4,x_3,x_2,x_5,x_8,x_7\}\neq \text{pos}_C(D)$$

由此可见，去掉属性 C_4 不改变决策表的分类能力，所以属性 C_4 为冗余属性，在决策表中去除 C_4，同时并没有出现相同的实例，故此为决策表的最后约简。如表 10-3 所示。

表 10-3　寄生虫案例最终决策表

U	C_1	C_2	C_3	C_5	D
x_1	0	1	1	0	0
x_2	0	1	0	1	0
x_3	1	1	0	0	1
x_4	1	1	1	0	1
x_5	1	0	0	1	0
x_6	0	1	1	1	1
x_7	0	0	1	1	1
x_8	1	1	0	1	1

10.2.4　BP 神经网络设计和算法

BP 多层前馈神经网络模型是有监督学习的典型网络，有监督学习是指神经网络在实际应用中解决各种问题时，必须对它进行训练，就是从应用环境中选出一些样本数据，通过不断调整权矩阵，直到得到合适的输入输出关系为止，这个过程就是网络的训练过程，这种训练过程需要提供训练数据。

(1) BP 神经网络的设计

本研究的神经网络获取步骤采用的是 BP 模型，虽然模型已广泛应用，但其设计却主要依靠经验来进行。

输入输出层设计：BP 网络输入输出层维数是完全根据实际使用情况决定的。在本模型中，根据鱼病的种类不同，输入数据的维数也略有不同，但输出维数均为 1，这样输出层的节点数就确定为 1。输入层输入的是对象信息，在本研究中即是相关鱼病的症状。比如对象“寄生虫”，其包含的信息是 5 个属性：吃食不好、病鱼逛游、头部发黑、腮部肿胀和腮部充血。则对于“寄生虫”这个对象，可以定义 5 个输入神经元。输出层输出的即是鱼病类型。

隐层设计：隐层的设计包括隐层层数和每层神经元个数的确定，主要依靠经验进行。对于层数的选择，隐含层起抽象作用，即它能从输入提取特征。增加隐含层可增加神经元网络的处理能力，本研究采用一个隐含层，因为经验表明对小型的网络边界判决问题，两层并不比单层更优越，同时鱼病领域知识问题属于二进制分类方面的问题，一个隐含层就已经足够了。

关于隐含层的神经元的数目，我们采用 Hecht-Nielsen 提出的算法来确定，即单层隐含层节点数目为 $2N+1$，其中 N 为输入的节点数。本研究通过约简后每个鱼病案例集的症状属性个数来确定隐含层神经元的数目。

(2) BP 神经网络算法

鱼病案例样本包括症状集合和疾病集合，经过向量化后，组成数值化的症状-疾病样本集。在学习过程中对整个训练样本集，根据神经网络的实际输出模式和期望输出模式的误差调整网络的权值，直到误差的均方值小于某一预定极小值，网络达到稳定。此时，神经网络就从这些数值化所表示的诊断经验中获取到了知识，并将这些知识分布存储在网络中。基本算法如下：

步骤 1：全部权值与节点的阈值预置为一个小随机值；

步骤 2：输入矢量 $\boldsymbol{X}$(鱼病特征属性)与输出矢量 $\boldsymbol{T}$(鱼病种类)；

步骤 3：计算实际输出矢量 $\boldsymbol{Y}$；

步骤 4：修整权值；

步骤 5：达到预定误差精度或循环次数后退出，否则转步骤 2。

10.3 知识获取实例

下面以鱼病病例库中的赤皮病诊断案例知识为例子，简单说明鱼病诊断案例知识获取的整个过程。首先将鱼病的特征属性用 C 表示，其中 C_1 表示厌食，C_2 表示死亡，C_3 表示鳃丝多黏液，C_4 表示鳃盖充血，C_5 表示鳞片脱落，C_6 表示体表出血，C_7 表示鳍出血。D 表示决策属性，值域设为(0=没有，1=有)，即用量化的包括条件属性和决策属性的决策表表示原始的赤皮病诊断案例表，如表 10-4 所示。

表 10-4 属性量化后建立的赤皮病决策表

U	C_1	C_2	C_3	C_4	C_5	C_6	C_7	D
X_1	0	1	1	0	1	1	0	1
X_2	1	0	0	1	0	0	1	0
X_3	1	0	1	1	0	1	0	1
X_4	1	0	0	1	1	0	1	1
X_5	0	0	1	1	0	1	0	0
X_6	1	0	0	1	1	0	1	1
X_7	0	1	0	0	1	0	1	1
X_8	0	1	1	0	1	0	0	0
X_9	1	1	1	0	1	0	1	1
X_{10}	0	1	0	0	1	1	0	1
X_{11}	0	1	1	0	1	0	0	0
X_{12}	0	1	1	1	1	0	0	1

删除决策表中重复的行 X_6，然后利用粗糙集理论对决策表进行约简，约简方法如前所述，主要是计算条件属性对决策属性的重要程度，通过计算 $\mathrm{pos}_C(D)$，判断出 C_2、C_7 是冗余的，可以从决策表中删除，从而得到属性约简后的决策表如表 10-5 所示。



表 10-5　属性约简后的决策表

U	C_1	C_3	C_4	C_5	C_6	D
X_1	0	1	0	1	1	1
X_2	1	0	1	0	0	0
X_3	1	1	1	0	1	1
X_4	1	0	1	1	0	1
X_5	0	1	1	0	1	0
X_6	0	0	0	1	0	1
X_7	0	1	0	1	0	0
X_8	1	1	0	1	0	1
X_9	0	0	0	1	1	1
X_{10}	0	1	0	1	0	0
X_{11}	0	1	1	1	0	1

利用约简后的属性组合，将 C_1、C_3、C_4、C_5、C_6 作为神经网络的输入，用 D 作为输出，输入输出样本表如表 10-6 所示，训练 BP 网络前设定网络层数为三层，输入层神经元个数为 5 个，隐含层和输出层的神经元个数分别为 10 个和 1 个。

表 10-6　基于粗糙集的 BP 神经网络的输入和输出样本

样本号 U	输入样本					输出样本 D
	C_1	C_3	C_4	C_5	C_6	
X_1	0	1	0	1	1	1
X_2	1	0	1	0	0	0
X_3	1	1	1	0	1	1
X_4	1	0	1	1	0	1
X_5	0	1	1	0	1	0
X_6	0	0	0	1	0	0
X_7	0	1	0	1	0	1
X_8	1	1	0	1	0	1
X_9	0	0	0	1	1	0
X_{10}	0	1	0	1	0	0
X_{11}	0	1	1	1	0	1

将表 10-6 的前 5 个实例作为网络的训练集进行训练，设定训练次数为 50 次，为保证学习的稳定性，这里取学习率为 0.03，当期望误差取 0.001 时，经过 8 步训练后，误差 MSE＝0.000 560 721＜0.001达到要求。实验中 BP 神经网络的训练是收敛的。

将表 10-6 中的后 6 个例子作为测试样本，计算结果如表 10-7 所示，实验证明测试结果与实际结果误差很小。

表 10-7　测试样本对照表

样本号 U	输入样本					计算结果 D'	输出样本 D
	C_1	C_3	C_4	C_5	C_6		
X_6	0	0	0	1	0	0.1650	0
X_7	0	1	0	1	0	0.9559	1
X_8	1	1	0	1	0	0.9862	1
X_9	0	0	0	1	1	0.1023	0
X_{10}	0	1	0	1	0	0.1104	0
X_{11}	0	1	1	1	0	0.9976	1

因此说明可以通过粗糙集和神经网络结合获取规则，以其中两个为例表示如下：

if C_1(厌食)/0 **and**
C_3(鳃丝多黏液)/0 **and**
C_4(鳃盖充血)/ 0 **and**
C_5(鳞片脱落)/1 **and**
C_6(体表出血)/0
then D(赤皮病)/0.1650

if C_1(厌食)/0 **and**
C_3(鳃丝多黏液)/1 **and**
C_4(鳃盖充血)/ 0 **and**
C_5(鳞片脱落)/1 **and**
C_6(体表出血)/0
then D(赤皮病)/0.9559

10.4　小结

本章通过对粗糙集理论和神经网络知识获取方法的文献分析，结合鱼病案例知识的特点，建立了基于粗糙集和神经网络松耦合的鱼病知识获取模型。研究结果表明，该模型很好地将两种方法进行优势互补，粗糙集作为前端处理机制，简化了神经网络结构，降低了神经网络原始特征空间的维数，提高了神经网络的收敛速度、训练精度和获取规则的准确性。

参考文献

1. 黎明，张化光. 基于粗糙集的神经网络建模方法研究[J]. 自动化学报，2002，28(1)
2. 何明，冯博琴，马兆丰. 基于粗糙集神经网络的网络故障诊断新方法[J]. 西安交通大学学报，2004，12：1241～1246
3. 何先华，赵卫东，盛昭瀚. 粗糙集在医疗诊断知识支持中的应用[J]. 计算机工程与应用，2001，20：166～168
4. 陈允平，王旭蕊，韩宝亮. 人工神经网络原理及其应用[M]. 北京：中国电力出版社，2002

第11章

基于遗传算法的知识获取模型

本章采用遗传算法和机器学习的理论方法，针对知识获取问题，在传统机器学习方法AQ的基础上，提出基于GA的知识获取模型，并应用于鱼病诊断推理，将鱼病诊断的知识获取问题转化为组合优化问题，实验结果表明该方法是有效的。

11.1 基于归纳的机器学习

11.1.1 机器学习的基本概念

1. 基本概念

机器学习是人工智能的一个重要研究领域，19世纪80年代以来，机器学习已被视为解决专家系统中知识获取“瓶颈”问题的关键。一般认为，机器学习是一个有特定目的的知识获取过程。[1]机器学习框架如图11-1所示，学习系统旨在根据提供的一组概念样本和背景知识，确定特定概念的描述。

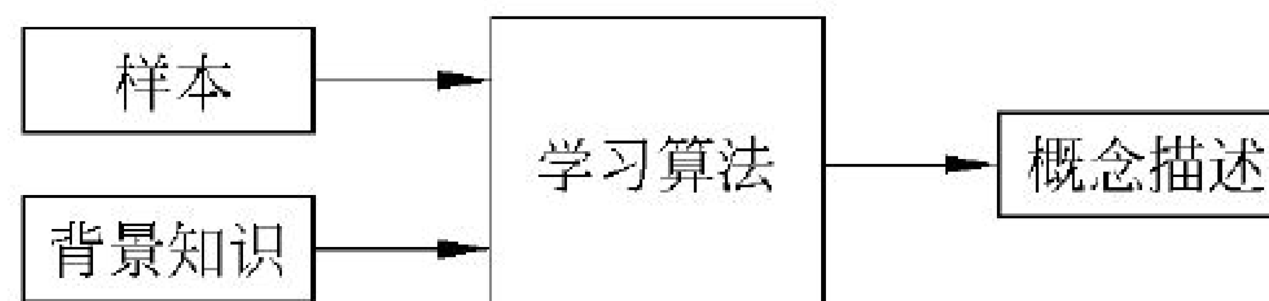


图11-1 机器学习框架

学习算法一般分为两大类别：黑箱法和基于知识的方法。黑箱法有自己的概念表示方法，然而其内部表示不易被用户所解读，且对其识别过程没有提供明确的说明或解释。黑箱法通常涉及相关参数、距离或权重的数值计算。基于知识的方法旨在创建满足可理解原则的符号知识结构。[2]我们根据鱼病诊断推理系统的实际情况，探讨基于知识的方法的机器学习算法。

概念可以理解为对一组具有某种共同性质而区别于其他对象的抽象表示。通常情况下，按照概念的功能将其分为陈述性概念和推理性概念。陈述性概念用于识别和定义一类具有显著物理含义的或发生频繁的事物，而推理性概念旨在反映事物之间的因果关系或者相关关系。陈述性概念是推理性概念分析的基础。

给定数据库中的一个实例集合，在机器学习上称为背景集，$E=E_1\times E_2\times\cdots\times E_n$ 是 n 维有限向量空间，其中 $E_j(j=1,2,\cdots,n)$ 为有限集合。E 中的某个实例 $e=\langle e_1,e_2,\cdots,e_n\rangle$ 称为例子，其中 $e_j\in E_j(j=1,2,\cdots,n)$。

将 E 分为正例集 PE 和反例集 NE：

$$PE=\{e_1^+,e_2^+,\cdots,e_{K_p}^+\},\quad e_i^+=\langle e_{i1}^+,e_{i2}^+,\cdots,e_{in}^+\rangle,\quad K_p<n$$

$$NE=\{e_1^-,e_2^-,\cdots,e_{K_n}^-\},\quad e_i^-=\langle e_{i1}^-,e_{i2}^-,\cdots,e_{in}^-\rangle,\quad K_n\leqslant n$$

假设 PE 和 NE 在 n 维属性集合 X 下是一致的，满足 $PE \cup NE = E$，$PE \cap NE = \varnothing$，$K_p + K_n = n$，$|PE| = K_p$，$|NE| = K_n$。

概念学习就是对于一个给定的背景集，构造蕴涵这些概念的归纳语言，选择最佳或满意的归纳语言，即可以解释观察实例的一组概念。

2. 学习的归纳本质

相关概念通常能组织成树状或图所表示的泛化层次结构。在一个泛化层次结构中，一个概念不仅可通过低层对象来说明，而且也可以通过求解概念以下任何一级的子概念来说明。例如：概念“烂鳃病”就是“鱼病”的一个实例，“赤皮病”也是一个实例（如图 11-2 所示）。

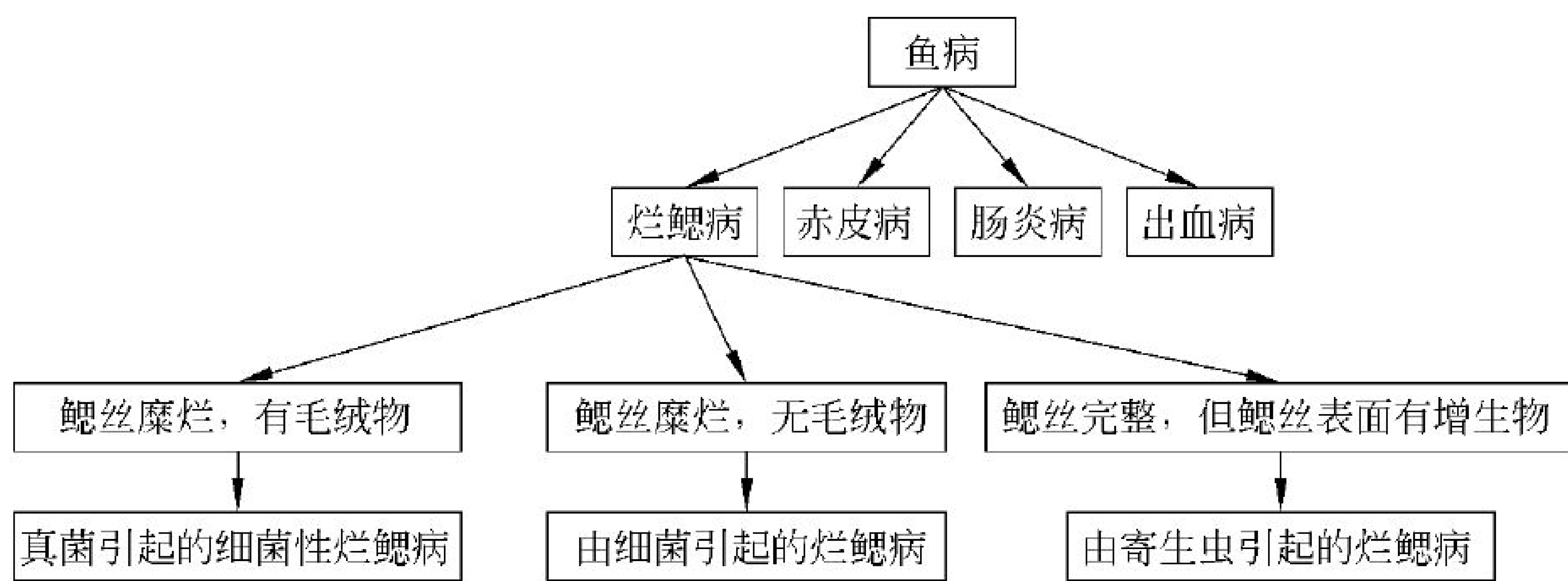


图 11-2　泛化层次结构

在学习“烂鳃病”概念的过程中，基本的学习策略首先给出烂鳃病的所有症状（头部乌黑、鳃盖开天窗）作为概念第一个正例，学习者简单记住该病的所有特征，在学习过第一个正例后，学习者仅知道此正例最特殊的描述而不知道其他特征，这个描述成为一组最特殊描述（用 S 表示）。此时对最特殊描述进行任何泛化都是可能的。为确定其局限性，学习者将需要一个反例样本，即给出一个白头白嘴病的症状（体色发黑、口部溃烂及嘴端至眼球皮肤变为乳白色），仅在第一个反例（白头白嘴病）加入后，学习者可能对泛化进行某种限制，因此获得一组最泛化描述（用 G 表示）来正确地覆盖正例。从这个集合 G 中，学习者选择“鳃盖开天窗”作为最吸引人的特征。

下一个反例赤皮病（鳃盖出血）表明描述过于泛化，鳃部不能区分烂鳃病（正例）和赤皮病（反例）。因此集合 G 必须进行细化。另一方面，下一个正例将用一个更特殊的描述丰富集合 S，从而需要通过把“鳃盖开天窗”替换为“鳃丝腐烂”和“鳃盖开天窗”来完成另一次泛化。

总而言之，当一个新正例样本加入时，就应用泛化来处理集合 S。相反，当一个反例加入集合 G 时，就进行细化。这个原则构成了一系列称为版本空间算法的技术基础。

3. 泛化空间的搜索

假设选择了某种表示语言，而且学习者想要从一组正例与反例样本中学习一个概念，采用的描述是基于属性-值逻辑的，其全部可识别概念的空间是非常大的，假设有 10 个属性，每个属性可取 5 个值，则将构成 $5^{10} = 9\,765\,625$ 个可能的向量。这些向量的任意子集都对应一个概念，即意味着根据这些属性可以定义 $2^{9\,765\,625}$ 个概念。在更复杂的语言中，这个数目甚至会增加得更快。



为了处理问题的可计算性,学习者大多结合归纳与启发搜索方法来进行计算。[3,4]

(1) 穷尽搜索

概念学习的框架就是搜索由学习者表示语言所容许的描述空间。一般而言,一个搜索过程将探寻搜索空间中的每个状态。

定义 11.1 初始状态:即搜索的起点,在机器学习中,初始状态常常对应最特殊的概念描述,也就是正例本身。

定义 11.2 终止条件:即要到达的目标。满足终止条件的状态称为最终状态。

定义 11.3 搜索操作符:用以实现从一个状态转到另一个状态的搜索,这种操作符大多是概念描述的泛化和/或细化。

定义 11.4 搜索策略:即确定在什么条件下,操作符可以用于什么样的状态。

系统搜索的两个基本策略:深度优先搜索和广度优先搜索。

(2) 启发式搜索

基本的搜索技术对于大搜索空间而言效率较低,通常要考虑启发式搜索,启发式任务就是确定哪些操作可使搜索接近最终状态。启发式搜索算法通常分为最好优先搜索算法和集束搜索算法。最好搜索算法与广度优先搜索不同,因为它总是仅扩展最有希望的状态,希望以此加速搜索。其代价是可能陷入评价函数局部最优的窘境。由于要存储产生的所有状态,最优搜索有时需要很大的存储空间,而集束搜索算法是一种比较经济的方法,它在任何时候仅保留 N 个最好状态。

11.1.2 覆盖正例排斥反例的 AQ 学习方法

AQ 学习是基于通过连续产生决策规则来主动覆盖训练数据的思想,该方法在国际上具有较大影响,已从最初由 Michalski 首先提出的一个算法,发展成为一系列算法。[4~6]其本质就是搜索一组规则以覆盖任意正例但不覆盖任一反例。AQ 算法不划分样本,而是对所选择的正例描述进行泛化。

1. AQ 方法的基本概念

定义 11.5 关系命题(selector):关系命题又叫选择子,形式为$[X_i \# A_j]$,其中 X_i 是例子 e 的第 j 个变元,$A_i \subseteq D_i$,关系运算符 $\# \in \{=,\neq,>,\geqslant,<,\leqslant\}$。

定义 11.6 复合(complex):关系命题的合取称为复合,其中的合取符号“∧”常常省略。

定义 11.7 覆盖(cover):覆盖是复合的析取。

定义 11.8 例子 e 满足公式:假设例子 $e=\langle V_1,V_2,\cdots,V_n\rangle$,关系命题 $S=[X_i \neq A_i]$及公式(复合)$L=\wedge[X_i \neq A_i]$,则

① e 满足选择子 S 当且仅当 $V_i \notin A$;

② e 满足 L 当且仅当 e 满足 L 中的每一个选择子,即对所有的 $j \in J$,$V_i \notin A_i$;

③ e 满足 S(或 L)也叫作 S(或 L)覆盖 e。

定义 11.9 星(star):正例 e^+ 对于反例集 NE($e^+ \notin$ NE)的星为 G,它是覆盖 e^+ 但不覆盖 NE 中任何一个反例的复合中最一般的复合。最一般是指不存在同样的复合 G' 使 $G \subset G'$。

在 AQ 方法中,Star 生成算法是生成一个正例 e^+ 对于反例集 NE 的星。

2. AQ方法的核心算法

AQ系列方法具有从示例中学习、渐进学习、构造式学习、测试等功能。其核心算法是Star生成算法和Cover生成算法。

Star生成算法是根据每个正例(种子)生成排斥(不覆盖)所有反例(反例集)的星集star。Cover生成算法是根据每个种子生成的星集star,再生成覆盖集Cover,即最后生成的覆盖cover能够覆盖所有正例并且排斥所有反例。

(1) Cover生成算法

步骤1:当“部分cover”已经覆盖所有的正例,则转步骤6;

步骤2:选择一个未被覆盖的正例作为种子(seed);

步骤3:调用Star生成算法,生成这个seed对反例集的星的集合;

步骤4:根据用户定义的标准,选择一个最佳的星(其实是一个复合);

步骤5:把这个星加入到“部分cover”中,生成一个新的“部分cover”,最后得到的“部分cover”就是所求的判定规则;

步骤6:结束。

(2) Star生成算法

步骤1:当“部分star”不覆盖任何反例时,则转步骤6。

步骤2:选择一个被覆盖的反例(从“部分star”中清除它,使该反例不被覆盖)。

步骤3:调用Cover CC算法,生成seed的“排斥反例的部分star”。它是这样一个覆盖cover,即覆盖seed,但排斥这个反例,它是简单覆盖,即选择子的析取。

步骤4:Multify LL算法对旧的“部分star”求交集,生成新的“部分star”。

步骤5:若新的“部分star”中的复合的个数超过了用户定义的个数,则按用户定义标准,去掉“质量差”的复合,最后得到的“部分star”作为算法的输出,即一个星的集合。

步骤6:结束。

(3) Cover CC算法

步骤1:对正例e^+中的选择子与相应的选择子的交集为空则转步骤2,否则去掉选择子;

步骤2:用选择子代替中的选择子,使其不覆盖当前反例相应属性的取值;

步骤3:最后得到选择子的析取,即为排斥当前反例的“部分star”;

步骤4:结束。

(4) Multify LL算法

设旧的“部分star”为cover1(它是由复合的析取组成),排斥当前反例的“部分star”为cover2(它是由选择子selector的析取组成的)。

对于cover1的每一个复合和cover2的每一个selector进行合取操作(即求交集),每次返回一个复合,最后返回的是复合的析取,即为所要求的新的“部分star”。由于新的“部分star”的复合会有冗余,需要进行简化,以去掉冗余的复合。

11.1.3 AQ方法的应用

示例学习就是从已知的正例和反例中,归纳出能够描述正例而排斥反例的一般规则。在学习的过程中,它既需要正例集又需要反例集。数据库中的元组集合可以被视为示例集合,我们可以将某一类元组作为正例集,其余类所有的元组作为反例集,这样依次指定正例集和反例



集,便可以发现知识表中支持一类元组而排斥其余类的一般规则,即分类规则。假设有烂鳃病和草鱼出血病,其中每一种病的症状参见表 11-1。

表 11-1　诊断病例的样本数据

ID	疾病	肌肉	体表	腹部	鳞片	头部	鱼鳍	鳃部	肠道
1.1	烂鳃病	000	000	000	000	B01	000	C01	000
1.2	烂鳃病	000	000	000	000	B01	000	C05	000
1.3	烂鳃病	000	000	000	000	B01	000	C06	000
2.1	草鱼出血病	G03	000	000	000	B02	F02	C03	000
2.2	草鱼出血病	G03	000	000	000	B03	F02	C03	000
2.3	草鱼出血病	000	A01	000	000	B02	F02	C03	000
2.4	草鱼出血病	000	A05	000	000	B02	F02	C03	000
2.5	草鱼出血病	000	A05	000	000	B03	F02	C03	000

由于上述两种鱼病的症状与腹部、鳞片、肠道无关,可以不予考虑。在鱼病诊断系统中,症状集合如表 11-2 所示。

表 11-2　鱼病症状代码表

A01	体色发黑	B06	口腔可见针形虫体
A02	体色苍白	B07	眼眶充血
A03	背鳍或臀鳍至尾鳍的体色呈白色	B08	眼球突出
A04	鱼体局部、全部出血发炎明显	B09	眼球混浊,呈乳白色,甚至眼球脱落
A05	鱼体两侧出血或充血	B10	脑廓呈黄色
A06	鱼体两侧有斑点状出血	B11	脑室及眼眶周围或脑室中央部位呈鲜红色
A07	背部皮肤出现脓疮,其内部充满血脓和大量细菌	C01	鳃盖开天窗
A08	尾部或腹部两侧有印章形红斑	C02	鳃盖张开
A09	体表产生隆起的白色斑块,呈石蜡状	C03	鳃盖出血
A10	体表形成一层白色黏液薄膜(头背、尾鳍明显)	C04	鳃丝黏液多
A11	体表有白色或淡蓝色黏液附着	C05	鳃丝腐烂
A12	体表粗糙	C06	鳃丝腐烂发白,附淤泥和杂物尖端软骨外露
A13	体表有点状白色黏液物	C07	鳃丝肿大
A14	体表有许多棉毛状菌丝,并附有污泥和藻类	C08	鳃瓣斑点或点状充血
A15	体表有白翳附着	C09	鳃瓣苍白(白鳃)
A16	体表有白色小点状胞囊	C10	鳃瓣粉红色
A17	体表有灰白色香肠形、带形或线形胞囊	C11	鳃瓣有白色小点状胞囊
A18	体表有许多黑色小结节	C12	鳃上有白色胞囊
A19	体表有透明圆形虫体	C13	鳃上有灰白色香肠形、带形或线形胞囊
A20	体表可见针形虫体	C14	鳃瓣上有白色蛆状虫体
A21	体表有白色胞囊	C15	有透明圆形虫体
B01	头部乌黑	C16	鳃上有椭圆形虫体并带有棒状乳白色卵囊
B02	头部充血	C17	鳃丝糜烂,有毛绒物
B03	口腔充血	C18	鳃丝糜烂,无毛绒物
B04	唇肿胀、口部溃烂、嘴端至眼球皮肤变乳白色	C19	鳃丝完整,但鳃丝表面有增生物
B05	口部溃烂、黏液分泌	D01	腹部膨大

续表

D02	腹水	G03	肌肉点状或块状出血
D03	腹部膨大，肛门红肿外突	G04	肌肉腐烂形成小坑甚至露出骨骼和内脏
D04	腹水，肛门红肿外突	H01	肠道充血不糜烂
E01	鳞片脱落	H02	肠道充血发炎
E02	腹部两侧表皮腐烂，病灶处鳞片明显脱落	H03	肠道糜烂
E03	鳞片张开似松球基部积水	H04	肠道无食、充满积液或血脓
E04	鳞片竖起，鳞下有线形虫体	H05	前肠粗大，肠壁呈白色糜烂状
E05	鳞片竖起，鳞下有胞囊	H06	肠道有胞囊
F01	蛀鳍	H07	肠道或性腺有白色圆形或椭圆形胞囊
F02	鳍基出血	H08	肠道穿孔，肠外壁形成结节
F03	鱼鳍上有胞囊	H09	肠壁有许多白色结节状病灶，肠管粗大
F04	鱼鳍上有白色小点状胞囊	H10	体腔内充满白色弯曲的裂头蚴虫体
F05	鳍条有许多黑色小结节	H11	肠道有带状虫体聚居
F06	尾鳍有白色镶边(严重时腐烂)	H12	鳔内壁有点状出血，肝、脾、肾水肿
G01	肌肉充血呈红色	H13	胆囊肿大，胆汁淡黄色或无色甚至外溢
G02	肌肉出血呈暗红色	H14	肝脏发紫，严重时萎缩

示例学习的目的就是要从两类(或多类)例子集中找出描述一类而排斥其余类的一般规则。我们从鱼病分类的例子来看 AQ 方法的执行过程。

初始的“部分 cover”为空，该类中有两个决策者，每个类中的例子都进行编号。

首先选择烂鳃病为当前类即正例集，草鱼出血病便成为反例集。

① 因为初始的“部分 star”为全体事件的空间，可以表示为

$$[肌肉 \neq \varnothing][体表 \neq \varnothing][头部 \neq \varnothing][鳃部 \neq \varnothing][鱼鳍 \neq \varnothing] \quad (1)$$

② 进入 Cover 生成算法，选择一个正例，烂鳃病 seed1.1，表示为

[肌肉＝000][体表＝000][头部＝B01][鳃部＝C01][鱼鳍＝000]

并对反例集排序得 2.3、2.1、2.2、2.4、2.5。

③ 进入 Star 生成算法，选择反例 2.3，表示为

[肌肉＝000][体表＝A01][头部＝B02][鳃部＝C03][鱼鳍＝F02]

调用 Cover CC 算法，计算排斥当前反例 2.3 的“部分 star”得

$$[体表\neq A01][头部\neq B02][鳃部\neq C03][鱼鳍\neq F02] \quad (2)$$

调用 Multify LL 算法得(1)和(2)的交集，即新的“部分 star(1)”为

CPX1：$[体表 \neq A01]\ [头部 \neq \varnothing]\ [鳃部 \neq \varnothing]\ [鱼鳍 \neq \varnothing]$

CPX2：$[体表 \neq \varnothing]\ [头部 \neq B02]\ [鳃部 \neq \varnothing]\ [鱼鳍 \neq \varnothing]$

CPX3：$[体表 \neq \varnothing]\ [头部 \neq \varnothing]\ [鳃部 \neq C03]\ [鱼鳍 \neq \varnothing]$

CPX4：$[体表 \neq \varnothing]\ [头部 \neq \varnothing]\ [鳃部 \neq \varnothing]\ [鱼鳍 \neq F02]$

④ 新的“部分 star(1)”覆盖反例 2.1，调用 Cover CC 算法，计算排斥当前反例 2.1 的“部分 star(2)”得

$$[肌肉\neq G03]\ [头部\neq B02][鳃部\neq C03]\ [鱼鳍\neq F02]$$

调用 Multify LL 算法求新的“部分 star(2)”与旧的“部分 star(1)”的交集，得

CPX1：$[体表 \neq A01]\ [头部 \neq B02]\ [鳃部 \neq \varnothing]\ [鱼鳍 \neq \varnothing]$

CPX2：$[体表 \neq \varnothing]\ [头部 \neq B02]\ [鳃部 \neq \varnothing]\ [鱼鳍 \neq \varnothing]$



CPX3：[体表 ≠ ∅][头部 ≠ B02][鳃部 ≠ C03][鱼鳍 ≠ ∅]
CPX4：[体表 ≠ ∅][头部 ≠ B02][鳃部 ≠ ∅][鱼鳍 ≠ F02]
CPX5：[体表 ≠ A01][头部 ≠ ∅][鳃部 ≠ C03][鱼鳍 ≠ ∅]
CPX6：[体表 ≠ ∅][头部 ≠ B02][鳃部 ≠ C03][鱼鳍 ≠ ∅]
CPX7：[体表 ≠ ∅][头部 ≠ ∅][鳃部 ≠ C03][鱼鳍 ≠ ∅]
CPX8：[体表 ≠ ∅][头部 ≠ ∅][鳃部 ≠ C03][鱼鳍 ≠ F02]
CPX9：[体表 ≠ A01][头部 ≠ ∅][鳃部 ≠ ∅][鱼鳍 ≠ F02]
CPX10：[体表 ≠ ∅][头部 ≠ B02][鳃部 ≠ ∅][鱼鳍 ≠ F02]
CPX11：[体表 ≠ ∅][头部 ≠ ∅][鳃部 ≠ C03][鱼鳍 ≠ F02]
CPX12：[体表 ≠ ∅][头部 ≠ ∅][鳃部 ≠ ∅][鱼鳍 ≠ F02]

简化上述 12 个复合，得“部分 star(3)”，即

CPX1：[体表≠∅][头部≠B02][鳃部≠∅][鱼鳍≠∅]
CPX2：[体表≠∅][头部≠∅][鳃部≠C03][鱼鳍≠∅]
CPX3：[体表≠∅][头部≠∅][鳃部≠∅][鱼鳍≠F02]

⑤“部分 star(3)”已覆盖反例 2.2，调用 Cover CC 算法，计算排斥当前反例 2.2 的“部分 star(4)”得

[肌肉≠ G03][头部≠B03][鳃部≠C03][鱼鳍≠F02]

调用 Multify LL 算法，求“部分 star (4)”与“部分 star(3)”的交集，得

CPX1：[体表≠∅][头部≠B02∧B03][鳃部≠∅][鱼鳍≠∅]
CPX2：[体表≠∅][头部≠B03][鳃部≠C03][鱼鳍≠∅]
CPX3：[体表≠∅][头部≠B03][鳃部≠∅][鱼鳍≠F02]
CPX4：[体表≠∅][头部≠B02][鳃部≠C03][鱼鳍≠∅]
CPX5：[体表≠∅][头部≠∅][鳃部≠C03][鱼鳍≠∅]
CPX6：[体表≠∅][头部≠∅][鳃部≠C03][鱼鳍≠F02]
CPX7：[体表≠∅][头部≠B02][鳃部≠∅][鱼鳍≠F02]
CPX8：[体表≠∅][头部≠∅][鳃部≠C03][鱼鳍≠F02]
CPX9：[体表≠∅][头部≠∅][鳃部≠∅][鱼鳍≠F02]

简化上述 9 个复合，可得“部分 star(5)”为

CPX1：[体表≠∅][头部≠B02∧B03][鳃部≠∅][鱼鳍≠∅]
CPX2：[体表≠∅][头部≠∅][鳃部≠C03][鱼鳍≠∅]
CPX3：[体表≠∅][头部≠∅][鳃部≠∅][鱼鳍≠F02]

⑥“部分 star(5)”覆盖反例 2.4 调用 Cover CC 算法，计算排斥当前反例 2.4 的“部分 star(6)”得

[体表≠ A05][头部≠B02][鳃部≠C03][鱼鳍≠F02]

调用 Multify LL 算法，求“部分 star(6)”与“部分 star(5)”的交集，得

CPX1：[体表 ≠ A05][头部 ≠ B02 ∧ B03][鳃部 ≠ ∅][鱼鳍 ≠ ∅]
CPX2：[体表 ≠ A05][头部 ≠ ∅][鳃部 ≠ C03][鱼鳍 ≠ ∅]
CPX3：[体表 ≠ A05][头部 ≠ ∅][鳃部 ≠ ∅][鱼鳍 ≠ F02]
CPX4：[体表 ≠ ∅][头部 ≠ B02 ∧ B03][鳃部 ≠ ∅][鱼鳍 ≠ ∅]
CPX5：[体表 ≠ ∅][头部 ≠ B02][鳃部 ≠ C03][鱼鳍 ≠ ∅]

CPX6：[体表 ≠ ∅] [头部 ≠ B02] [鳃部 ≠ ∅] [鱼鳍 ≠ F02]

CPX7：[体表 ≠ ∅] [头部 ≠ B02 ∧ B03] [鳃部 ≠ C03] [鱼鳍 ≠ ∅]

CPX8：[体表 ≠ ∅] [头部 ≠ ∅] [鳃部 ≠ C03] [鱼鳍 ≠ ∅]

CPX9：[体表 ≠ ∅] [头部 ≠ ∅] [鳃部 ≠ C03] [鱼鳍 ≠ F02]

CPX10：[体表 ≠ ∅] [头部 ≠ B02 ∧ B03] [鳃部 ≠ ∅] [鱼鳍 ≠ F02]

CPX11：[体表 ≠ ∅] [头部 ≠ ∅] [鳃部 ≠ C03] [鱼鳍 ≠ F02]

CPX12：[体表 ≠ ∅] [头部 ≠ ∅] [鳃部 ≠ ∅] [鱼鳍 ≠ F02]

简化上述 12 个复合，可得

CPX1：[体表≠∅] [头部≠B02∧B03] [鳃部≠∅] [鱼鳍≠∅]

CPX2：[体表≠∅] [头部≠∅] [鳃部≠C03] [鱼鳍≠∅]

CPX3：[体表≠∅] [头部≠∅] [鳃部≠∅] [鱼鳍≠F02]

⑦ “部分 star(6)”覆盖反例 2.5，调用 Cover CC 算法，计算排斥当前反例 2.5 的“部分 star(7)”，得

[体表≠ A05] [头部≠B03][鳃部≠C03] [鱼鳍≠F02]

调用 Multify LL 算法，求“部分 star(7)”与“部分 star(6)”的交集，得

CPX1：[体表≠A05] [头部≠B02∧B03] [鳃部≠∅] [鱼鳍≠∅]

CPX2：[体表≠A05] [头部≠∅] [鳃部≠C03] [鱼鳍≠∅]

CPX3：[体表≠A05] [头部≠∅] [鳃部≠∅] [鱼鳍≠F02]

CPX4：[体表≠∅] [头部≠B02∧B03] [鳃部≠∅] [鱼鳍≠∅]

CPX5：[体表≠∅] [头部≠ B03] [鳃部≠C03] [鱼鳍≠∅]

CPX6：[体表≠∅] [头部≠ B03] [鳃部≠∅] [鱼鳍≠F02]

CPX7：[体表≠∅] [头部≠B02∧B03] [鳃部≠C03] [鱼鳍≠∅]

CPX8：[体表≠∅] [头部≠∅] [鳃部≠C03] [鱼鳍≠∅]

CPX9：[体表≠∅] [头部≠∅] [鳃部≠C03] [鱼鳍≠F02]

CPX10：[体表≠∅] [头部≠B02∧B03] [鳃部≠∅] [鱼鳍≠F02]

CPX11：[体表≠∅] [头部≠∅] [鳃部≠C03] [鱼鳍≠F02]

CPX12：[体表≠∅] [头部≠∅] [鳃部≠∅] [鱼鳍≠F02]

简化上述 12 个复合，可得

CPX1：[体表 ≠ ∅] [头部 ≠ B02 ∧ B03] [鳃部 ≠ ∅] [鱼鳍 ≠ ∅]

CPX2：[体表 ≠ ∅] [头部 ≠ ∅] [鳃部 ≠ C03] [鱼鳍 ≠ ∅]

CPX3：[体表 ≠ ∅] [头部 ≠ ∅] [鳃部 ≠ ∅] [鱼鳍 ≠ F02]

又因为[体表≠∅]包含了属性“体表”的整个值域，所以应当去掉这个选择子，得到“部分 star(8)”如下：

CPX1：[头部≠B02∧B03] [鳃部≠∅] [鱼鳍≠∅]

CPX2：[头部≠∅] [鳃部≠C03] [鱼鳍≠∅]

CPX3：[头部≠∅] [鳃部≠∅] [鱼鳍≠F02]

⑧ 此时，“部分 star(8)”已经不覆盖所有的反例了，将它加入到“部分 cover”中，又因为“部分 star(8)”已经覆盖全部正例，所以 Cover 生成算法结束，得到第一组的判定规则如下：

CPX1：[头部 ≠ B02 ∧ B03] [鳃部 ≠ ∅] [鱼鳍 ≠ ∅]

CPX2：[头部 ≠ ∅] [鳃部 ≠ C03] [鱼鳍 ≠ ∅]



CPX3：[头部 ≠ ∅][鳃部 ≠ ∅][鱼鳍 ≠ F02]

其他两组以此类推。

11.2　遗传算法在知识获取中的应用

知识系统的核心就是知识，知识的数量和质量是决定一个系统性能优劣的主要因素。在知识系统的开发研究中，知识获取占用的人力与时间最多，是领域专家一致认定的“瓶颈”问题。知识获取研究的最终目标是构造一个具有一定自治能力、自动的知识获取系统，包括机器感知、机器识别和机器学习。将遗传算法用于知识获取，可以构成以 GA 为核心的机器学习系统。

遗传算法是一种基于生物自然选择与遗传机理的随机搜索算法。其基本思想是 Darwin 的进化论和 Mendle 的遗传学说。该算法由美国密执安大学教授 Holland 及其学生于 1975 年创建。作为一种随机的优化与搜索方法，遗传算法与传统算法相比有以下鲜明的特点。

(1) 遗传算法操作不直接作用在解空间上，它的可行解集利用解的某种编码表示(通常采用二进制编码)，不受函数约束条件的限制，因而具有良好的可操作性和简单性。

(2) 遗传算法的操作对象是一组可行解，而非单个可行解。即从一个群体上进行寻优，可以避免搜索过程陷入局部最优解，这是它能以较大的概率找到整体最优解的主要原因，因而具有良好的并行性。

(3) 遗传算法虽然只对 N 个个体操作，但实际上处理了大约 $O(n^3)$ 个模式，因而遗传算法具有隐含的并行性，有助于处理非线性问题。

(4) 遗传算法只使用解的适应性信息进行工作，并在增加收益和减小开销之间进行权衡，适用于大规模、高度非线性的、不连续的多峰函数的优化，具有很强的通用性。

(5) 遗传算法引用了概率转移规则，能搜索离散的、带有噪声的、多峰值的复杂空间，因而具有很强的鲁棒性。

(6) 遗传算法采用随机操作，采用“优胜劣汰、适者生存”原理激励结构向最优解前进，它的方向性使其效率远远高于一般的随机算法。

11.2.1　概念获取模型

示例学习的目标即是寻找最优覆盖规则，即生成具有最少数目的覆盖，这已被证明为 NP 难题。利用遗传算法来寻求最优覆盖规则是一个新的尝试。根据 AQ 算法的思想，结合鱼病诊断知识，我们按如下方法构造模型：第一类约束条件要求第一个正例必须被至少一个原子公式及所代表的逻辑规则所覆盖，第二类约束条件要求除第一个正例之外的全部正例必然被公式所覆盖，第三类约束条件要求全部反例不能被任何公式所覆盖，或者只能被该公式的某些原子公式所覆盖。该模型将概念学习问题转化为一个规划模型。

本模型中，根据实例集合 E，采用 CNF 范式规则表示要获取的概念 C，要求概念 C 满足完整性和一致性，并且尽量简单或概念的长度越短越好。

若 $\forall e_i^+ \in \mathrm{PE}(i=1,2,\cdots,K_\mathrm{P})$，$F(e_i^+,g)=1$，称 $F(g)$ 满足完整性；

若 $\forall e_i^- \in \mathrm{NE}(i=1,2,\cdots,K_\mathrm{N})$，$F(e_i^-,g)=0$，称 $F(g)$ 满足一致性。

以概念 C 满足完整性和一致性为约束条件，以简单性为目标，建立如下整数规划模型：

$$\left.\begin{array}{ll}\text{Min } Z = \sum_{j=1}^{n}\sum_{l=1}^{m_j} x_{jl} + \omega \sum_{j=1}^{n} x_j & \\ \sum_{j=1}^{n} x_{1l_j j} \geqslant 1 & \\ \sum_{j=1}^{n} x_{il_j j} - \sum_{j=1}^{n} x_{1l_j j} < 0, & i = 1,2,\cdots,K_{\mathrm{P}} \\ \sum_{j=1}^{n} x_{il_j j} - \sum_{j=1}^{n} x_{1l_j j} < 0, & i = 1,2,\cdots,K_{\mathrm{N}} \\ x_j = x_{j1} \vee x_{j2} \vee \cdots \vee x_{jm_j}, & j = 1,2,\cdots,n \end{array}\right\} \tag{11-1}$$

其中 $x_{il_j j}=x_{jl}(l=1,2,\cdots,m)$。$x_j=x_{j1}\vee x_{j2}\vee\cdots\vee x_{jm_j}$ 表示若 $\forall l, x_{jl}=0$,则 $x_j=0$;若 $\exists\, l, x_{jl}=1$,则 $x_j=1$。ω 是一个比较大的惩罚系数,一般 $\omega\geqslant 2\times\max\{m_j\}(j=1,2,\cdots,n)$。

11.2.2 基于遗传算法的规则学习实例

以草鱼出血病和烂鳃病两种病例规则为例,病例特征和症状如表 11-3 与表 11-4 所示。以草鱼出血病作为正例,烂鳃病作为反例,说明基于 GA(遗传算法)的概念学习过程和特点。在实际中,我们得到的实例集合的属性变量和值域如表 11-5 所示。

表 11-3 草鱼出血病(红鳍红鳃盖型)诊断病例的样本数据

ID	疾 病	肌肉	体表	腹部	鳞片	头部	鱼鳍	鳃部	肠道
1	草鱼出血病(红鳍红鳃盖型)	G03	000	000	000	B02	F02	C03	000
2	草鱼出血病(红鳍红鳃盖型)	G03	000	000	000	B03	F02	C03	000
3	草鱼出血病(红鳍红鳃盖型)	G03	000	000	000	B07	F02	C03	000
4	草鱼出血病(红鳍红鳃盖型)	000	A05	000	000	B02	F02	C03	000
5	草鱼出血病(红鳍红鳃盖型)	000	A05	000	000	B03	F02	C03	000
6	草鱼出血病(红鳍红鳃盖型)	000	A05	000	000	B07	F02	C03	000
7	草鱼出血病(红鳍红鳃盖型)	000	A01	000	000	B02	F02	C03	000
8	草鱼出血病(红鳍红鳃盖型)	000	A01	000	000	B03	F02	C03	000
9	草鱼出血病(红鳍红鳃盖型)	000	A01	000	000	B07	F02	C03	000

表 11-4 烂鳃病诊断病例的样本数据

ID	疾病名称	肌肉	体表	腹部	鳞片	头部	鱼鳍	鳃部	肠道
1	烂鳃病	000	000	000	000	B01	000	C01	000
2	烂鳃病	000	000	000	000	B01	000	C05	000
3	烂鳃病	000	000	000	000	B01	000	C06	000

表 11-5 现有疾病实例集合的属性和值域

序号	x_{ij}	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
		肌肉	体表	腹部	鳞片	头部	鱼鳍	鳃部	肠道
1	x_{i0}	G03	000	000	000	B02	F02	C03	000
2	x_{i1}	000	A05	000	000	B01	F02	C01	000
3	x_{i2}	G03	000	000	000	B07	000	C03	000
4	x_{i3}	000	000	000	000	B03	000	C05	000
5	x_{i4}	000	A01	000	000	B03	F02	C06	000



根据式(11-1),这个问题的规则模型为

$$
\begin{aligned}
\min Z = & x_{10}+x_{12} \\
& +x_{20}+x_{21}+x_{25} \\
& +x_{50}+x_{51}+x_{52}+x_{53}+x_{54} \\
& +x_{60}+x_{61}+x_{64} \\
& +x_{70}+x_{71}+x_{72}+x_{73}+x_{74} \\
& +\omega(x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8)
\end{aligned}
$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}\geqslant 1$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{10}-x_{12}-x_{53}-x_{54}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{10}-x_{12}-x_{52}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{21}-x_{50}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{21}-x_{53}-x_{54}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{21}-x_{52}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{24}-x_{50}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{24}-x_{53}-x_{54}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{24}-x_{52}-x_{60}+x_{61}-x_{64}-x_{70}-x_{72}=0$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{51}-x_{71}\geqslant 1$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{51}-x_{73}\geqslant 1$$

$$x_{10}+x_{12}+x_{50}+x_{60}+x_{61}+x_{64}+x_{70}+x_{72}-x_{51}-x_{74}\geqslant 1$$

$$x_1=x_{10}\vee x_{12}$$

$$x_2=x_{21}\vee x_{24}$$

$$x_5=x_{50}\vee x_{52}\vee x_{53}\vee x_{54}$$

$$x_6=x_{60}\vee x_{61}\vee x_{64}$$

$$x_7=x_{70}\vee x_{72}$$

采用基于 GA 的规则学习步骤如下。

(1) 染色体的构造

对于给定例子集合 E,其属性集合 $X=\{x_1,x_2,\cdots,x_n\}$,第 j 个属性 x_j 的取值范围 $D_j=\{d_{j1},d_{j2},\cdots,d_{jm_j}\}$。采用 $x_j=\{x_{j1},x_{j2},\cdots,x_{jm_j}\}$($x_{jl}=0$ 或 1)表示 D_j 上的子集 g_j 的构造。那么 D 上整个子集组 G 的基于字符集{0,1}的编码位串表示为

$$s_i=x_{11}x_{12}\cdots x_{1m_1}x_{21}x_{22}\cdots x_{2m_2}\cdots x_{j1}x_{j2}\cdots x_{jm_j}\cdots x_{n1}x_{n2}\cdots x_{nm_n}$$

其中,位串的长度为 $L=\sum_{j=1}^{n}m_j$。

(2) GA 参数设置与算子设计

① max _gen:最大迭代次数;pop_size:群体大小;l_chrom:染色体长度;p_c:交叉概率;p_m:变异概率;ω:评价函数中的参数;gen=0。

② 群体初始化

```
for i=1 to pop_size do
for j=1 to l_chrom do
        染色体 Si 的第 j 位等位基因=rnd(low,high)
```

```
endfor
endfor
```

③ 评价函数

$$\text{Min } Z = \sum_{j=1}^{n}\sum_{l=1}^{m_j} x_{jl} + \omega\sum_{j=1}^{n} x_j$$

④ 选择算子

采用基于上述评价函数的选择机制，每次仅选择一个个体。

```
partsum=0
随机地产生一个实数 random∈[0,1]
for i=1 to pop_size do
    if(partsum<random) then
        选择 S_i;
    endif
endfor
```

⑤ 交叉算子

采用单点交叉方法进行染色体交叉。

```
for i=1 to pop_size/2 do
    随机地产生一个实数 random∈[0,1]
    if(random<p_c)then
        int j= rnd(pop_size);
        int k= rnd(pop_size);
        单点交叉产生两个新个体 S'_j,S'_k;
    endif
endfor
```

⑥ 变异算子

```
for i=1 to pop_size do
    随机地产生一个实数 random∈[0,1]
    if(random<p_m) then
        随机地产生需要变异的基因位置 j=rnd(0,l_chrome);
        对第 j 位进行变异;
    endif
endfor
```

⑦ 终止条件

```
if(gen<max _gen)then;
    令 gen=gen+1;
    转到评价函数;
else
    停止计算,输出结果。
endif
```



11.2.3　基于遗传算法的约束规划求解

用遗传算法求解约束优化问题有两种策略：一是群体中所有个体均满足给定的约束，并且要求遗传操作后生成的子代也满足给定的约束；二是群体中的个体可以在一定程度上违反给定的约束，但必须在个体的适应度计算上体现出约束的违反程度，通常这是通过惩罚函数来实现的。

遗传算法中惩罚函数法的基本思想是从传统优化中借鉴来的。遗传算法从包含大量个体的初始种群开始搜索最优解，通过选择算子把具有较好适应值的个体选择出来，然后进行交叉、变异，扩大搜索空间，形成下一代种群。对于约束优化问题，惩罚函数法对于任一违反了的约束，把一个惩罚项加入进化函数，使违反约束的个体的适应值降低，再通过选择算子，生成下一代种群，从而在群体中保持一定数量的非可行解，使遗传算法从可行域和不可行域两个方向进行搜索，找到全局最优解。

11.2.4　遗传算法与惩罚函数法混合的通用算法

以惩罚函数外点法为例，对于鱼病知识获取模型这样的不等式约束优化问题，惩罚函数外点法的求解步骤如下。

(1) 构造惩罚函数，将原问题转化为无约束的优化问题

$$\Phi(X,M^{(k)}) = f(X) + M^{(k)}\sum_{\alpha=1}^{m}\{\max[g\alpha(x),0]\}^{\alpha}$$

该式中的右边第二项为惩罚项；α 为构造惩罚函数的指数，其值将影响函数 $\Phi(X,M^{(k)})$ 等值线在约束面处的性质，一般取 $\alpha=2$；M 为惩罚因子，它是大于 0 的一个递增数列，即应满足

$$0 < M^{(0)} < M^{(1)} < \cdots < M^{(k)} < M^{(k+1)} < \cdots$$

$$\lim_{k\to\infty} M^{(k)} = +\infty$$

上式中 $M^{(0)}$ 为初始惩罚因子，可根据经验取值(比如取 $M^{(0)}=1$)。

(2) 适应度函数定义为

$$f(X) = C_0 - \Phi(X,M^{(k)})$$

此式中的 C_0 是一给定数，用以保证 $f(X)$ 为非负，这样就将求优化问题转化为求适应度最大的问题。

在这里将个体适应度进行某些转换可得到个体的选择几率，但可能会使得某些异常个体的选择几率过大。排序选择机制可以很好地控制个体的选择几率。

对群体中的所有个体按照适应值排序的机制如下：首先比较个体的约束型适应度 F_{com}，好的适应度的个体排名靠前，如果适应度值相等，则再比较其优化型适应度 F_{opt}，好的适应度排名靠前。

与通常的基于惩罚的方法相比，这样做使得可行点的适应度总优于不可行点，从而使得优化过程为先得到可行点，然后由这些可行点及较好的不可行点的遗传操作得到最优可行点。这样可以将进入可行区和得到优化点统一起来，且无须变换优化目标的适应度到大于零。由于不必设置 F_{com} 和 F_{opt} 的权重，使用起来较为简单。

排序后第 i 个个体的生存几率为

$$\text{prob}(i) = q(1-q)^{i-1}$$

其中 $q\in(0,1)$ 称为选择压力，用来控制个体被选择的几率，通常为平均个体选择几率的数倍。

为了保护弱小子代个体参与繁殖的能力，并防止某些异常个体在演化中被过多地选作父代使得基因趋同而导致早熟收敛，增加一个计数器 pNum 记录个体被选作父代的次数，第 i 个个体的生存几率为

$$\text{prob}(i)=q(1-q)^{i-1}(f_n)^{\text{pNum}}$$

其中 $q\in(0,1)$，可用来控制个体被选作父代的次数，pNum 越大，prob(i)几率越小。

(3) 标准基因遗传算法(CGA)常用于处理以下静态优化问题

$$\max\{f(b):b\in \text{IB}^L\}$$

其中：$0\leqslant f(b)\leqslant\infty$，$b\in \text{IB}^L=(0,1)^L$，$f(b)$不等于常量。

抽象地讲，CGA 由 7 个部分组成：

$$\text{CGA}=(\lambda,L,P^0,P',S,C,M)$$

其中：

$\lambda\in N$，为群体中的个体总数；

$L\in N$，为二进制编码串的长度；

$P^0=(b_1^0,b_2^0,\cdots,b_\lambda^0)\in I^\lambda$，$I=\text{IB}^L=(0,1)^L$；

$P'=(P_c,P_m)$，其中 P_c 为交叉概率，P_m 为变异概率；

S：$I^2\rightarrow I$ 为选择算子；

C：$I\times I^{P_c}$，$I\times I$ 为交叉算子。

一般 P_c 和 P_m 由使用者给定。

(4) CGA 的程序流程

```
初始化 L,λ,Pc,Pm 等;
随机产生初始群体,P旧;
while 终止条件不满足 do
    P新 的个体数量 N=0;
        for j=1 to λ do fj=f(bj)计算个体的适应度;
    最优解=max(fj)对应的个体;
        while N<λ do
          用选择算子 S 从 P旧 中选择双亲;
        if Pc≥Random(0,1) then 用交叉算子 C 产生两个后代;
        if Pm≥Random(0,1) then 用变异算子 M 改变该两个后代;
        把两个后代存入 P新
        N= N+2;
    EndWhile;
    P新 = P旧 ;
EndWhile;
输出最优值。
```

(5) 终止准则

可以使用三种终止准则：①利用本代的平均满足度值与上一代平均满足度值的比值作为终止准则；②利用最大循环次数作为终止准则；③利用群体最优基因个体多代未变作为终止准则。以上各种准则既可单独使用，也可联合使用。

通过 GA 求解可得出如下结论：

if (肌肉＝G03)∧(头部＝B03)∧(鱼鳍＝F02)∧(鳃部＝C03) **then** 草鱼出血病

if (肌肉＝G03)∧(体表＝A01)∧(鱼鳍＝F02)∧(鳃部＝C03) **then** 草鱼出血病



与实际的情况相符合，由此可以看出，该方法是有效的。

11.3　小结

知识获取是构造专家系统的“瓶颈”问题，而专家知识的好坏直接影响整个系统的性能。本章从机器学习的角度出发，研究了主动覆盖的 AQ 算法，并将其用于鱼病诊断的规则学习，并总结了 AQ 算法的优缺点。

由于鱼病诊断知识的学习实质上是一个组合问题，本章在 AQ 算法的基础上，根据 AQ 算法的基本思想与学习策略，建立了一个基于整数规划的概念获取模型，将知识获取问题转化为组合优化问题，该模型成功应用于鱼病诊断知识的获取。

遗传算法作为模拟自然选择与进化过程的一种随机化搜索算法，具有很强的鲁棒性。本章提出了将遗传算法用于求解概念获取模型，并证明该方法对鱼病诊断知识的获取是有效的。

参考文献

1. 恽为民，席裕庚. 遗传算法的运行机理分析[J]. 控制理论与应用，1996，13(3)：289～297
2. 张晓缋，戴冠中等. 遗传算法种群多样性的分析研究[J]. 控制理论与应用，1998，15(1)：17～23
3. Larranaga P, et al. Learning Bayesian networks structures by searching for the best ordering with genetic algorithms[J]. IEEE Transactions on System. Man and Cybernetics—Part A: System and Humans. 1996, 26(4): 487～493
4. Suzuki J. A Markov chain analysis on simple genetic aogorithms[J]. IEEE Transactions on System. Man and Cybernetics. 1995, 25(4): 655～659
5. Suzuki J. A further result on the Markov chain model of genetic algorithms and its application to a simulated annealing-like strategy[J]. IEEE Transactions on System. Man and Cybernetics—Part B: Cybernetics. 1998, 28(1): 95～102
6. Srinivas M, Patnaik LM. Adaptive probability of crossover and mutation in genetic algorithms[J]. IEEE Transactions on System. Man and Cybernetics. 1994, 24(4): 656～667
7. 马晓岩，倪骏. 一种改进型遗传算法及其收敛性分析[J]. 系统工程与电子技术，2000，22(9)：58～60
8. 何琳，王科俊等. 最优保留遗传算法及其收敛分析[J]. 控制与决策，2000，15(1)：63～66
9. 江瑞，罗予等. 一种协调勘探和开采的遗传算法：收敛性及性能分析[J]. 计算机学报，2001，24(12)：1233～1241
10. 徐宗本，聂赞坎等. 遗传算法的几乎必然强收敛性——鞅方法[J]. 计算机学报，2002，25(8)：785～793
11. 郭东伟，刘大有等. 遗传算法收敛性的动力学分析及其应用[J]. 计算机研究与发展，2002，39(2)：225～230
12. 涂承媛，涂承宇. 一种新的收敛于全局最优解的遗传算法[J]. 信息与控制，2001，30(2)：116～119
13. 王斌，李元香. 一种防止遗传算法过早收敛的“两阶段交替”算法[J]. 小型微型计算机系统，2003，24(3)：537～539
14. 喻寿益，郭观七. 遗传算法的平均收敛速度及其估计[J]. 控制理论与应用，2003，20(3)：467～469
15. 张讲社，徐宗本等. 整体退火遗传算法及其收敛充要条件[J]. 中国科学(E 辑)，1997，27(2)：154～164
16. 王霞，周国标. 整体退火遗传算法的几乎处处强收敛性[J]. 应用数学，2003，16(3)：1～7

第 12 章

基于径向基网络的梨黑星病中短期预测预警

梨树病虫害预警预测系统是复杂非线性的多因子综合系统，具有模糊、不确定等特性，适合采用结构简单、具有非线性拟合能力的径向基神经网络。本章在介绍人工神经网络和 BP 神经网络的基础上，分析径向基网络机理、学习过程，研究梨黑星病径向基网络预测模型的设计过程。从实测资料分析梨黑星病的时序变化，确定梨黑星病的中短期警情指标，构建梨黑星病病叶率、病果率、果实病的中短期预警模型，通过参数选择优化算法，并将病叶率径向基网络模型的预测结果与 BP 神经网络进行对比分析。

12.1 人工神经网络及 BP 神经网络

12.1.1 人工神经网络及其基本功能

人工神经网络是模仿生物神经系统的分布式存储、并行处理、自适应学习等特性而产生的一种具有大量连接的并行分布式处理器，它具有通过学习获取知识并解决问题的能力，且知识是分布存储在连接权中。自 1943 年 Mc Culloch 和 Pits 提出神经元的数学模型(简称为 M-P)后，人工神经网络取得了巨大的发展，成为一门具有独特风格的信息处理学科。这些发展表现在出现了各种人工神经网络、多种训练算法，并且在很多领域得到广泛应用。[1]

人工神经网络的一个节点如图 12-1 所示，其中，$x_1,x_2,\cdots,x_m$ 为节点的输入分量，它们是神经元所接收到的信息；$w_1,w_2,\cdots,w_m$ 为权重，也称连接强度；θ 为阈值，$R_i(x)$ 为激活函数，一般取为非线性函数；y 为节点的输出。

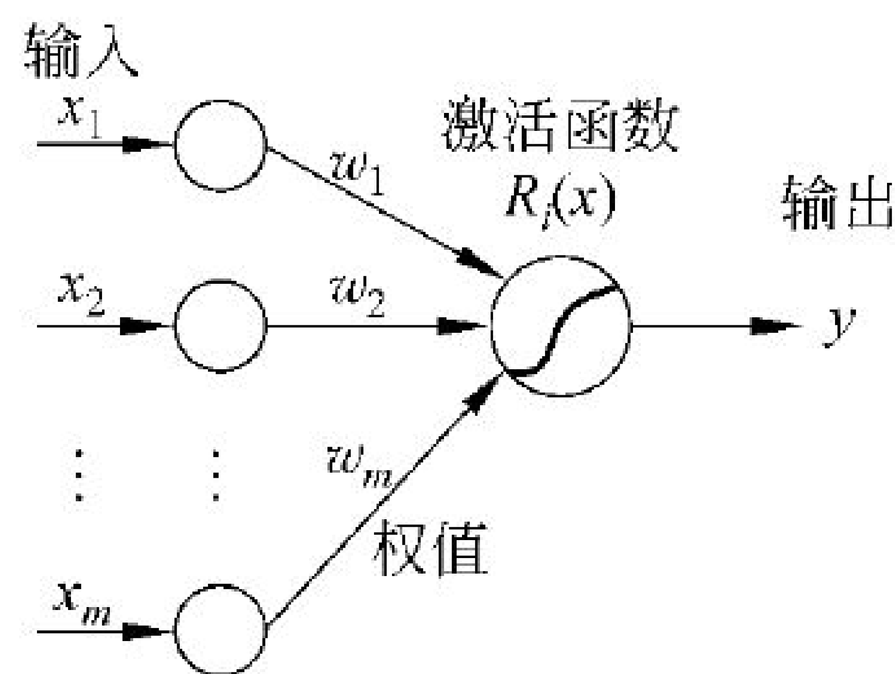


图 12-1 人工神经网络结构简图

人工神经网络的工作过程可分为训练和测试两个阶段。在训练阶段，以一组输入输出模式对作为训练样本集来训练网络，网络训练的过程即网络参数(包括权值、阈值等)的调整过程。在测试运行阶段，给定新的输入，网络即能计算得到相应的输出。

神经网络按照其基本模式分类，有前馈型、反馈型、自组织型、随机型网络等。在前馈型网络中主要有 Adaline、BP 神经网络、径向基网络(RBF 网络)，反馈网络主要有 Hopfield 网络，自组织网络主要有 ART 网络，随机网络主要有 Boltzman 机。神经网络按照其学习(或训练)方式可分为两种，一种是有监督或称有导师的学习，利用给定的样本标准进行分类或模仿；另一种是无监督学习或称无导师学习，只规定学习方式或某些规则，而具体的学习内容随系统所处的环境(即输入信号情况)而异，系统可以自动发现环境特征和规律性，具有更近似于人脑的



功能。人工神经网络分类如表 12-1 所示。

表 12-1 神经网络分类表

分类方式	神经网络分类
网络信息流向	前馈型网络、反馈型网络、混合网络
学习方式	有监督学习、无监督学习
连接方式	线性关联网络、非线性关联网络
应用对象	模式识别、预测预报、方案优化、智能控制、智能决策
算法不同	BP 网络、RBF 网络、Hopfield 网络、Boltzmann 机、Gauss 机

12.1.2 BP 神经网络

BP 网络反向传播算法是目前研究和应用较多的一类模型。BP 网络是一种多层前馈型网络，由输入层、输出层、中间层组成，中间层也称隐含层，其运算过程是一个映射，训练过程是动态的。BP 网络运行时，输入模式首先通过加权传到隐含层节点，经过激活函数的作用再传到输出节点。

BP 算法是一种有导师(有指导)学习算法，网络通过反向传播算法自动学习内部表达，也就是训练过程按误差由输出层经隐含层向输入层反向传播。整个训练过程为：首先根据当前的内部表达，对样本输入模式作前向计算，然后比较网络的输出与期望输出之间的误差，如误差小于规定值，则训练结束；否则，将误差向后传播，逐步调整权值和阈值，直至误差达到要求。因此，BP 算法又称误差反传算法。

BP 网络作为一种前馈型神经网络，不仅结构简单、易于实现，而且在许多方面表现出了优良的性能。目前已成为用输入输出数据模拟非线性模型的学习手段，其应用已渗透到几乎所有的工程领域，如自适应模式识别、实时语音重建、图像重构与数据压缩、实时分类、航空航天及机器人领域的大自由度系统控制，病虫发生程度的预测等。[2~4]

BP 神经网络的缺点也非常明显，它对于干扰非常敏感，训练过程不固定，另外，BP 网络用于函数逼近时，权值的调整采用梯度下降法，而且是采用全局逼近方法，即对每个输入输出数据对，网络的每一个权值均需调整，这些势必导致 BP 网络存在局部极小而得不到全局最优、训练时间长、收敛速度慢等缺点，影响预测效果。尽管现在对 BP 网络的改进方法很多，但都是片面地改善某一方面的性能。[5]

用径向基函数作神经元的功能函数构成的径向基神经网络，在一定程度上克服了 BP 网络的缺点，避免了 BP 网络繁琐、冗长的计算，在逼近能力、分类能力和学习速度等方面均优于 BP 网络，理论上为多层前馈网络的学习提供了一种新颖而有效的手段，因此，径向基网络的研究与应用越来越得到重视，并且广泛应用于数据分类和时间序列预测等方面。[6]这里采用径向基网络对梨黑星病的病叶率、病果率和果实病进行中短期预测。

12.2 径向基神经网络算法

12.2.1 径向基网络机理

1. 径向基网络结构

径向基网络是以径向基函数作为隐含层神经元激活函数的三层前馈型神经网络，其拓扑

结构如图 12-2 所示。第一层为输入层，由信号源节点组成；第二层为隐含层，其单元数由所描述问题的需要而定；第三层为输出层(以单输出为例)，它对输入模式的作用作出响应。

径向基网络工作的基本思想是：用径向基函数作为隐含层单元的“基”构成隐含层空间，它是一种局部分布的关于中心点对称的非线性函数，当径向基函数的中心点确定以后，就可以将输入矢量直接(而不通过权连接)映射到隐含层空间，映射关系随之确定。而隐含层空间到输出空间的映射是线性的，即网络的输出是隐单元输出的线性加权和。

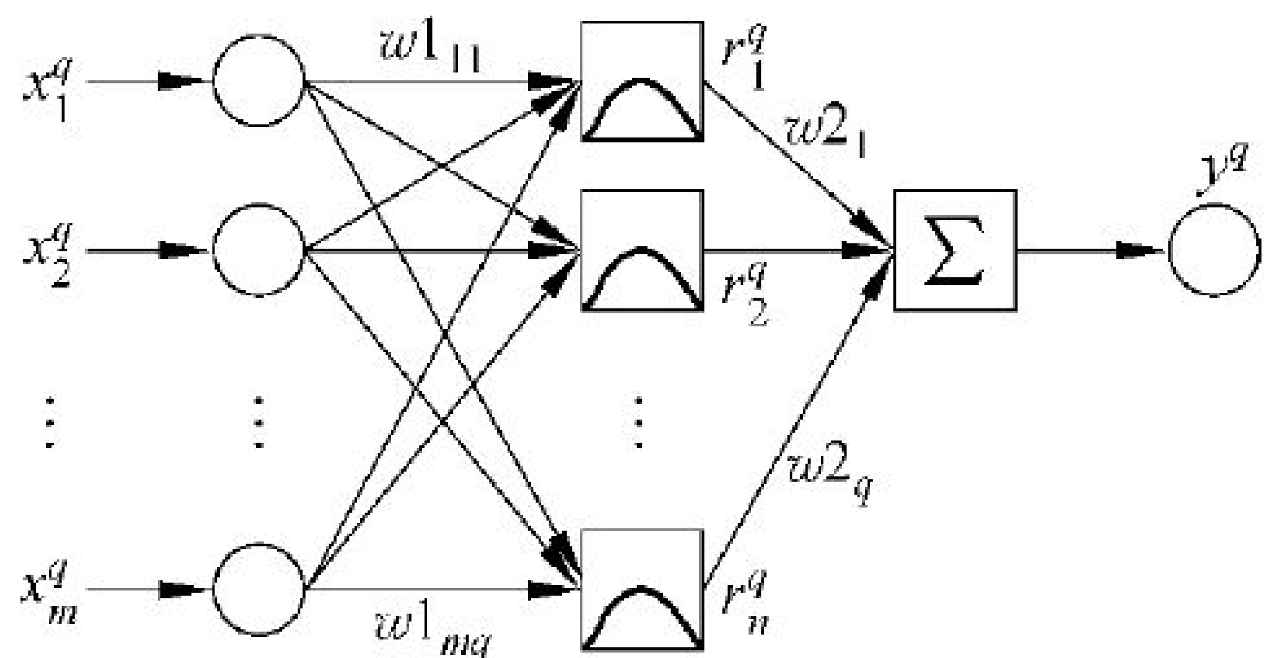


图 12-2　径向基网络结构

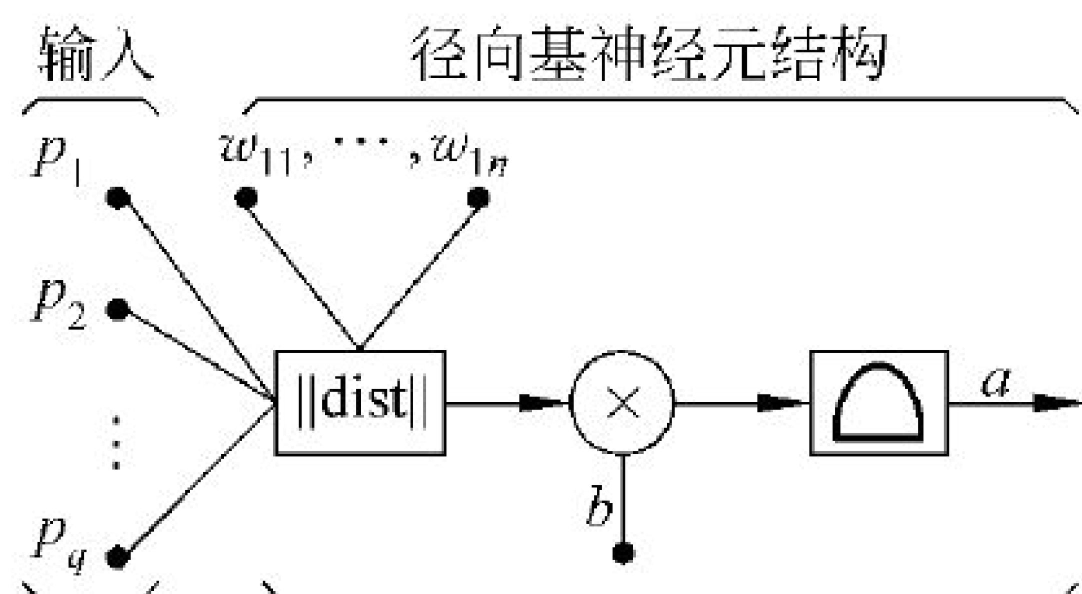


图 12-3　径向基网络神经元模型结构

2. 径向基神经网络的映射

径向基网络神经元模型结构如图 12-3 所示。径向基网络传递函数 radbas 以权值向量和阈值向量之间的距离 ‖dist‖ 为自变量，‖dist‖ 通过输入向量和加权矩阵的行向量的乘积得到，图中 b 为阈值，用于调整神经元的灵敏度。

径向基传递函数为

$$R_i(x)=\Phi\left[-\frac{\|\boldsymbol{x}-\boldsymbol{c}_i\|^2}{2\sigma^2}\right],\quad i=1,2,\cdots,m \tag{12-1}$$

式(12-1)中 $\boldsymbol{x}$ 是 m 维输入向量；$\boldsymbol{c}_i$ 是第 i 个基函数的中心，与 $\boldsymbol{x}$ 具有相同维数的向量；σ 是第 i 个感知的变量(可以自由选择的参数)，它决定了该基函数围绕中心的宽度，m 是感知单元个数。$\|\boldsymbol{x}-\boldsymbol{c}_i\|$ 是向量 $\boldsymbol{x}-\boldsymbol{c}_i$ 的范数，它通常表示 $\boldsymbol{x}$ 和 $\boldsymbol{c}_i$ 之间的距离，$R_i(x)$ 在 $\boldsymbol{c}_i$ 处有一个唯一的最大值，随着 $\|\boldsymbol{x}-\boldsymbol{c}_i\|$ 的增大，$R_i(\boldsymbol{x})$ 迅速衰减到零，对于给定的输入 $\boldsymbol{x}\in R_m$，只有一小部分靠近 $\boldsymbol{x}$ 的中心被激活。

从图 12-1 可以看出，输入层实现从 $\boldsymbol{x}\rightarrow R_i(x)$ 的线性映射，输出层实现从 $R_i(x)$ 到 $\boldsymbol{y}_k$ 的线性映射。即

$$\boldsymbol{y}_k=\sum w_{ik}R_i(\boldsymbol{x}),\quad k=1,2,\cdots,q \tag{12-2}$$

其中 q 是输出节点数。

3. 径向基函数

径向基函数的一个通用表达式为：

$$R(x)=\Phi((\boldsymbol{x}-\boldsymbol{c})^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{x}-\boldsymbol{c})) \tag{12-3}$$

其中：Φ 表示径向函数，$\boldsymbol{c}$ 表示函数的中心向量，$\boldsymbol{E}$ 是一个变换矩阵，通常为 Euclidean 矩阵。$(\boldsymbol{x}-\boldsymbol{c})^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{x}-\boldsymbol{c})$是在矩阵 E 定义的意义之下对输入向量 x 与中心 c 的距离的一种衡量。

如果 $\boldsymbol{E}$ 代表一个 Euclidean 矩阵，在这种情况下，$\boldsymbol{E}=\sigma^2\boldsymbol{I}$，$\sigma$ 为径向基函数半径，则上式简化为：

$$R(x)=\Phi((\boldsymbol{x}-\boldsymbol{c})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{c})/\sigma^2) \tag{12-4}$$



一般情况下，进一步简化为：

$$R(x) = \Phi\left(\frac{\| \boldsymbol{x} - \boldsymbol{c} \|^2}{\sigma^2}\right) \tag{12-5}$$

径向基函数还具有一个重要特征：随着与某一中心点距离的增大，函数曲线呈单调递减（递增）趋势。

下面是几类常用的径向基函数。

(1) 高斯函数(Gaussian function)

$$f(x) = \exp\left(-\frac{x^2}{\sigma^2}\right) \tag{12-6}$$

或

$$f(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{12-7}$$

(2) 薄板样条函数(thin-plate-spline function)

$$f(x) = x^2\log(x) \tag{12-8}$$

(3) 多二次函数(multiquadric function)

$$f(x) = (x^2 + \sigma^2)^{\frac{1}{2}} \tag{12-9}$$

(4) 逆二次函数(inverse multiquadric function)

$$f(x) = (x^2 + \sigma^2)^{-\frac{1}{2}} \tag{12-10}$$

径向基函数是局部分布且关于中心点径向对称衰减的非负非线性函数。它有两个主要参数：一个是基的中心，即对称点；另一个是基的宽度，即在多大的区域内会产生明显的输出响应。最常用的径向基函数是高斯函数。

12.2.2　径向基函数的学习过程

径向基神经网络是三层前馈神经网络，隐含层和输出层完成不同的功能。径向基网络逼近于非线性系统，在给定了学习数据后，算法主要解决结构设计和权值修正问题。结构设计包括如何确定网络隐含层节点数，如何确定径向基的聚类中心 c_j；权值修正就是调整隐含层空间到输出空间的权值矩阵 $\boldsymbol{W}$。

径向基网络输出层和隐含层所完成的任务不同，隐含层是对径向基的参数进行调整，采用的是非线性优化策略，输出层是对线性权进行调整，采用的是线性优化策略。因而，它们的学习策略也不同。径向基网络学习算法可以分为中心确定和权值调整两部分，二者可异步进行，也可同步进行，视具体问题的复杂度以及是否要求在线处理而定。

1. 径向基网络中心的确定

在设计径向基网络时，径向基中心的选择对网络性能的影响至关重要，中心选取不当，就不能设计出优化的径向基网络。例如，如果某些中心靠得太近，会产生近似线性相关，从而带来数值上的病变条件。所以确定径向基中心时，合理的选择并且尽量优化网络结构（使隐含层单元数尽可能少）是主要目的。目前，确定中心的方法大概可以分为以下几种。

(1) 随机选择径向基中心（直接计算法）

径向基中心是从输入样本数据中随机选取，且中心固定。当径向基中心确定以后，隐单元的输出是已知的，网络的连接权通过解线性方程组来确定。如果样本数据的分布具有代表性，

则此方法是简单可行的一种方法。但是大多数情况下，输入数据样本具有一定的冗余性时，这种算法就显得无能为力。

（2）根据经验选择中心

根据样本的分布情况，均匀选择若干个中心。

（3）自组织学习算法选取径向基中心

在这种方法中，径向基中心可以移动，并通过自组织学习确定其中心位置，而输出层的线性权则可以通过有监督学习规则计算。因此，这是一种混合的学习方法。自组织学习部分在某种意义上是对网络资源进行分配，学习的目的是使径向基网络的中心位于输入空间重要的区域，使选取的中心形成一个特定的分布规律，它表征着输入样本空间的固有特征。

（4）有监督学习选取径向基中心

在这种方法中，径向基中心以及网络的其他自由参数都是通过有监督的学习方法来确定的，通常通过非线性优化方法如梯度下降法、共轭梯度法等方法来确定。

（5）正交回归方法选取径向基中心

这是一种重要的径向基网络学习方法，径向基中心按照一定规则从样本数据中合理地选取，隐层单元数目在学习过程中动态调节，并且可以保证学习误差不大于给定值。

（6）支持向量方法自动选取径向基网络中心

在这种算法中，先用支持向量机算法获得支持向量，然后用支持向量作为径向基网络的中心，支持向量所起的作用就是获取径向基函数的中心，这样可以提高径向基网络的泛化能力。

（7）最近邻聚类学习算法

最近邻聚类学习算法是一种在线自适应动态聚类学习算法，不需要事先确定隐单元个数，完成聚类所得到的径向基网络是最优的，具有学习时间短、计算量小等特点，该算法具体过程如下：

① 选择一个适当的高斯函数宽度 r，定义一个矢量 $\mathbf{A}(l)$ 用于存放属于各类的输出矢量之和，定义一个计数器 $B(l)$ 用于统计属于各类的样本个数，其中 l 为类别数。

② 从第 1 个数据对 (x^1, y^1) 开始，在 x^1 上建立一个聚类中心，令 $c_1 = x^1$，$\mathbf{A}(1) = y^1$，$B(1)=1$，这样建立的径向基网络，只有一个隐单元，该隐单元的中心为 c_1，该隐单元到输出层的权矢量为 $\boldsymbol{w}_1 = \mathbf{A}(1)/B(1)$。

③ 考虑第 2 个样本数据对 (x^2, y^2)，求出 x^2 到 c_1 这个聚类中心的距离 $|x^2 - c_1|$，则：

如果 $|x^2 - c_1| \leqslant r$，则 c_1 为 x^2 的最近邻聚类，且令 $\mathbf{A}(1) = y^1 + y^2$，$B(1)=2$，$\boldsymbol{w}_1 = \mathbf{A}(1)/B(1)$；

如果 $|x^2 - c_1| > r$，则将 x^2 作为一个新聚类中心，并令 $c_2 = x^2$，$\mathbf{A}(2) = y^2$，$B(2)=1$。

在上述建立的径向基网络中再添加一个隐单元，该隐单元到输出层的权矢量为 $\boldsymbol{w}_2 = \mathbf{A}(2)/B(2)$。

④ 假设考虑第 k 个样本数据对 (x^k, y^k)，$k=3,4,\cdots,n$ 时存在 m 个聚类中心，其中心点分别为 $c_1, c_2, \cdots, c_m$。在上述建立的径向基网络中已有 m 个隐单元，再分别求出到这 m 个聚类中心的距离 $|x^k - c_j|$，$j=1,2,\cdots,m$。设 $|x^k - c_j|$ 为这些距离中的最小距离，即 c_j 为 x^k 的最近邻聚类，则

如果 $|x^k - c_j| > r$，则将 x^k 作为一个新聚类中心，并令 $c_{m+1} = x^k$，$m=m+1$，$\mathbf{A}(m) = y^k$，$B(m)=1$，且保持 $\mathbf{A}(i)$、$B(i)$ 的值不变 $(i=1,2,\cdots,m-1)$。在上述建立的径向基网络中再添加第 m 个隐单元，该隐单元到输出层的权矢量为 $\boldsymbol{w}_m = \mathbf{A}(m)/B(m)$。



如果$|x^k-c_j|\leqslant r$，作如下计算：$\boldsymbol{A}(j)=\boldsymbol{A}(j)+y^k$，$B(j)=B(j)+1$。当$i\neq j$时，$i=1,2,\cdots,m$，且保持$\boldsymbol{A}(i)$、$B(i)$的值不变。隐单元到输出层的权矢量为$\boldsymbol{w}=\boldsymbol{A}(i)/B(i)$。

⑤ 根据上述规则建立的径向基网络，其输出应为

$$f(x^k)=\frac{\sum_{i=1}^{m}w_i\exp(-\parallel x^k-c_i\parallel^2/r^2)}{\sum_{i=1}^{m}\exp(-\parallel x^k-c_i\parallel^2/r^2)}\tag{12-11}$$

在最近邻聚类学习算法中，高斯函数的宽度r是需要预先指定的。宽度r的大小决定了动态自适应径向基网络的复杂程度。r越小，所得到的聚类数目就越多，计算量就越大，精度也就越高；r越大，所得到的聚类数目就越少，计算量就越小，但精度也就越低。

由于r是一个一维参数，通常只需要根据误差信息来调整宽度r即可完成学习任务，这比同时确定隐单元个数和一个适合的范数要方便得多。由于每一个输入、输出数据对都可能产生一个新的聚类，因此，这种动态自适应径向基网络，实际上同时在进行参数和结构两个过程的自适应调整。

2. 径向基网络权值的调整

一般情况下，如果已知网络的隐层节点数和聚类中心，径向基网络从输入到输出就成了一个线性方程组。此时，权值学习可采用最小二乘法(least square，LS)求解。

只要径向基网络的聚类中心点和权值矩阵$\boldsymbol{W}$已知，则网络输入输出关系可由式(12-2)表示出来，这种方法简单易懂。由于其运算过程涉及矩阵求逆，对具有大量学习样本数据的系统，计算量无疑是很大的，所以径向基网络经常采用递推算法调整权值。常用的方法是递推最小二乘法(recursive linear least square，RLS)，此外常用的递推算法还有 Givens 最小二乘法、逆 Givens 最小二乘法。

上述方法为实际进行径向基网络设计提供了理论指导，可针对具体情况采用适当的方法。

12.3　梨黑星病径向基预测模型设计

12.3.1　梨黑星病数据准备

这里以安徽萧县梨黑星病为例，以萧县 1998—2006 年(1—12 月)的逐旬降水量、逐旬平均气温、逐旬平均湿度、逐旬累计雨日(≥0.1 mm)、逐旬日照时数为气象资料；采用萧县梨黑星病的测报资料：1999—2006 年(4—9 月)每 10 日的梨黑星病病叶率、病果率、病情发生指数等测报资料。

为避免输入向量物理意义和单位不同对径向基网络建模的影响，以便于进行计算和比较分析，对输入向量作归一化处理。

12.3.2　梨黑星病发生的实测时间动态

1. 病叶发生的时序变化

1999—2006 年梨黑星病病叶率的时序变化如图 12-4 所示。

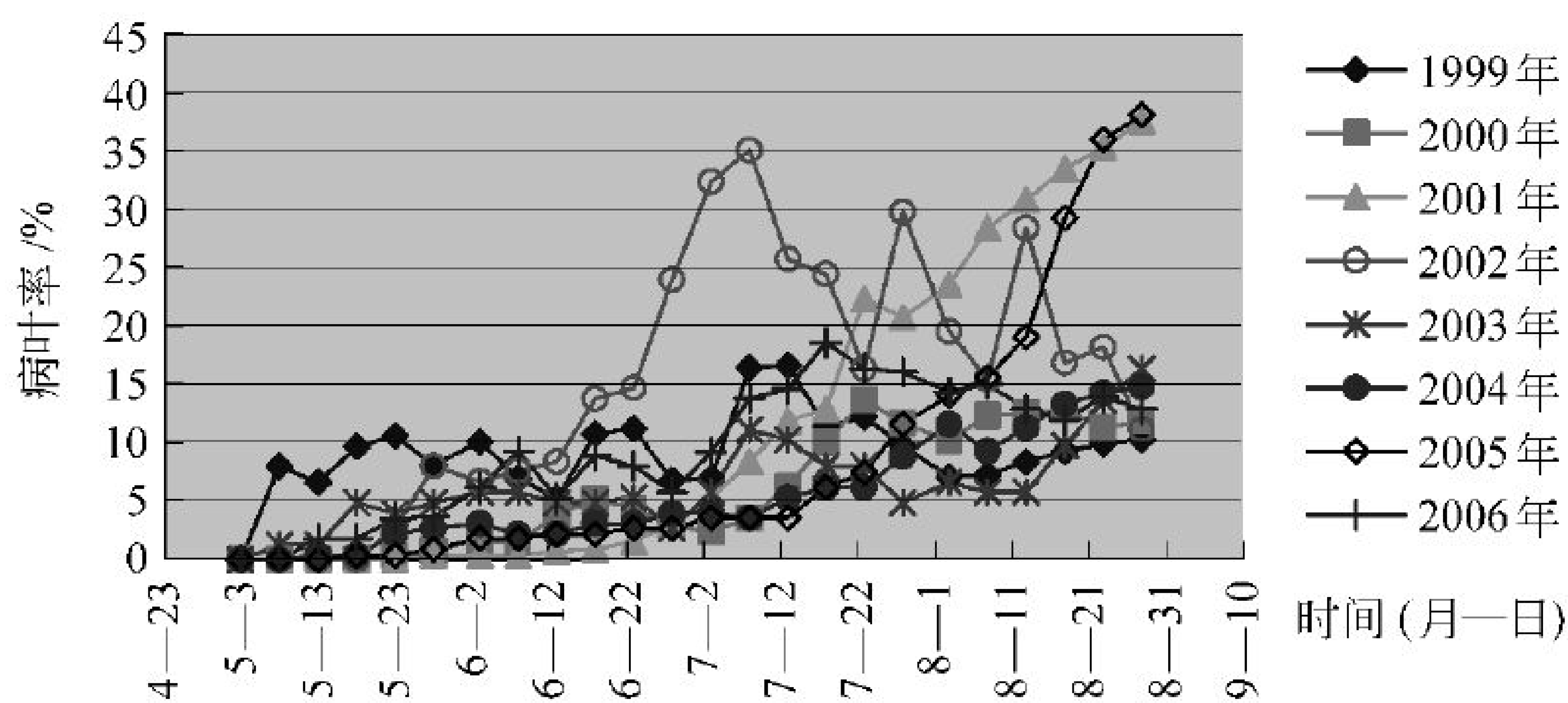


图 12-4　梨黑星病病叶率动态分析

2. 病果发生的时序变化

1999—2006 年梨黑星病病果率随时间的时序变化如图 12-5 所示。

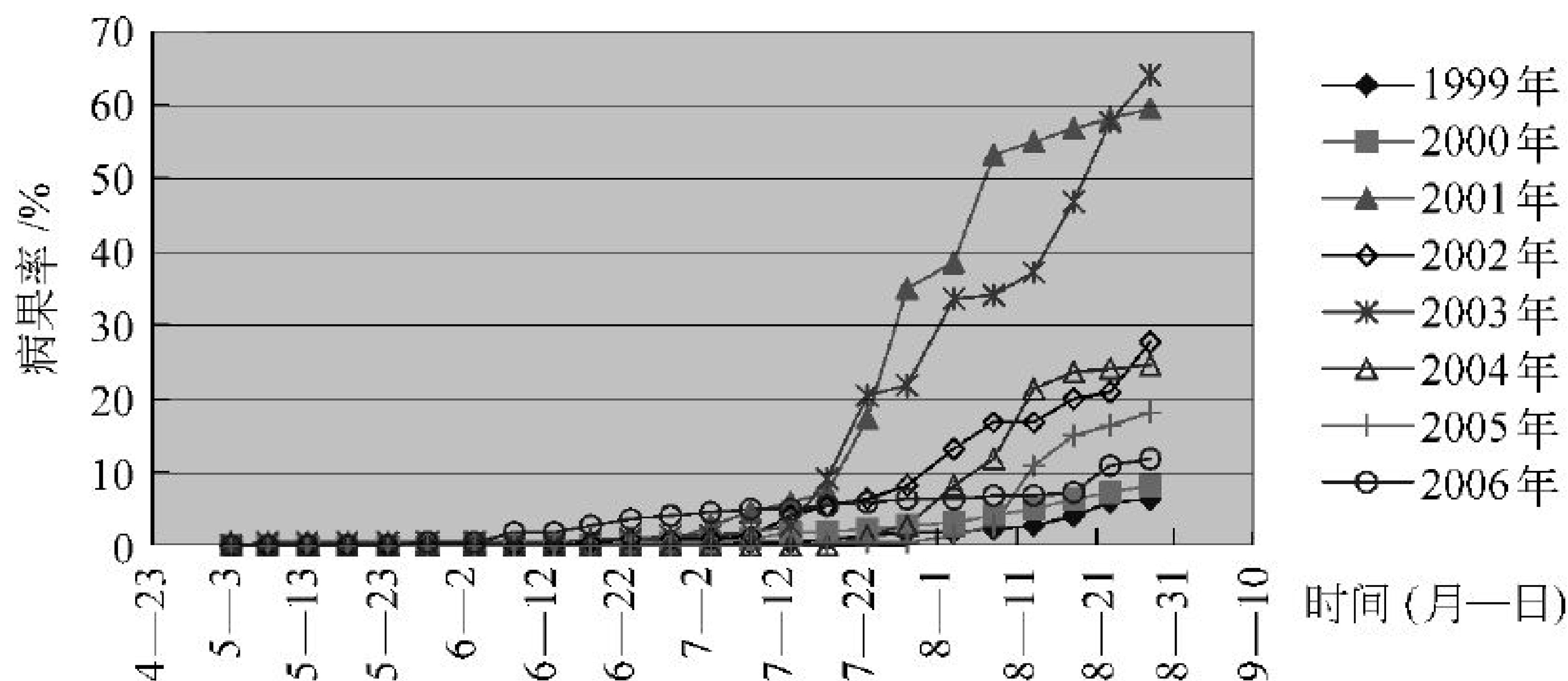


图 12-5　梨黑星病病果率动态分析

3. 果实病指的时序变化

1999—2006 年梨黑星病果实病指的时序变化如图 12-6 所示。

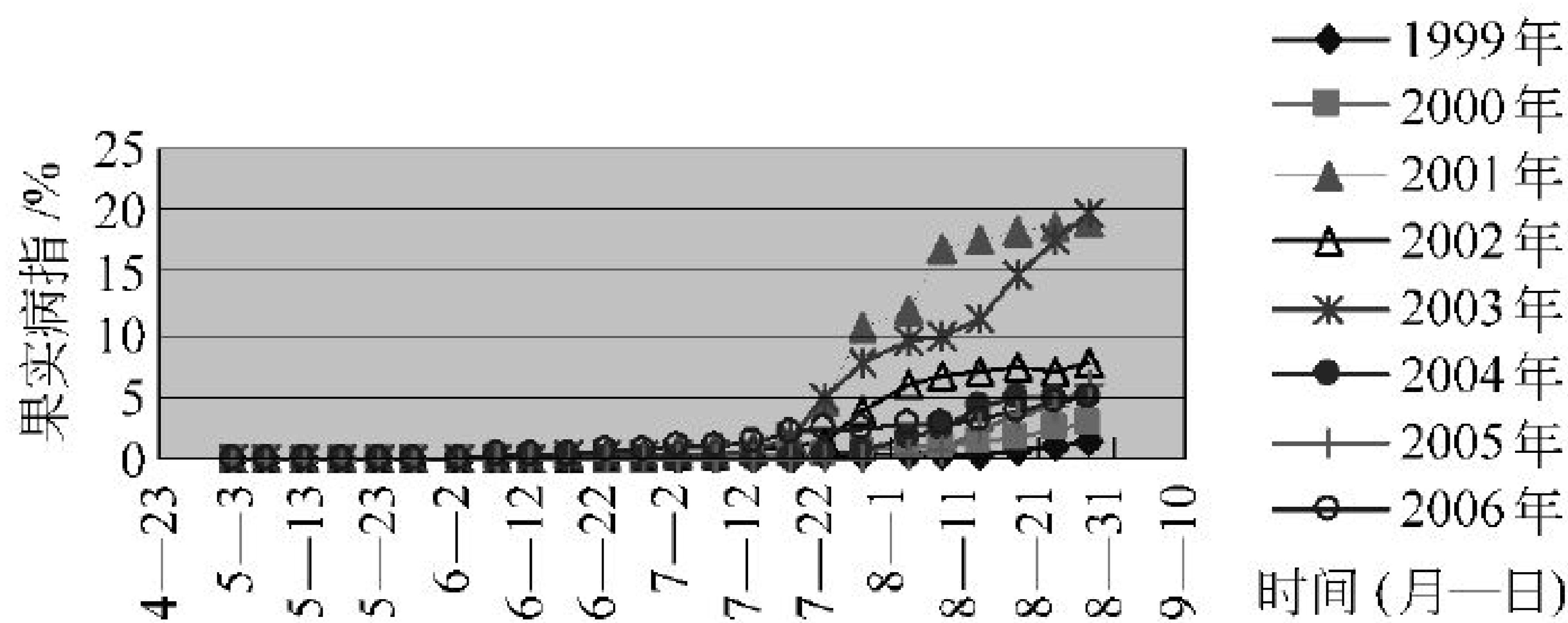


图 12-6　梨黑星病果实病指动态分析

12.3.3　梨黑星病径向基神经网络建模试验分析

1. 梨黑星病径向基神经网络结构

梨黑星病预测的径向基网络结构如图 12-7 所示，针对不同的病情指标预测，输入向量及输出向量各有不同，其中 $x_i^q(i=1,2,\cdots,m)$ 表示输入层单元，$w1_{iq}(i=1,2,\cdots,m)$ 表示输入层到隐藏层的权值，$w2_{iq}(i=1,2,\cdots,m)$ 表示隐藏层到输出层的权值，$r_i^q(i=1,2,\cdots,m)$ 表示第 i



单元的激励函数，y^q 表示输出值。

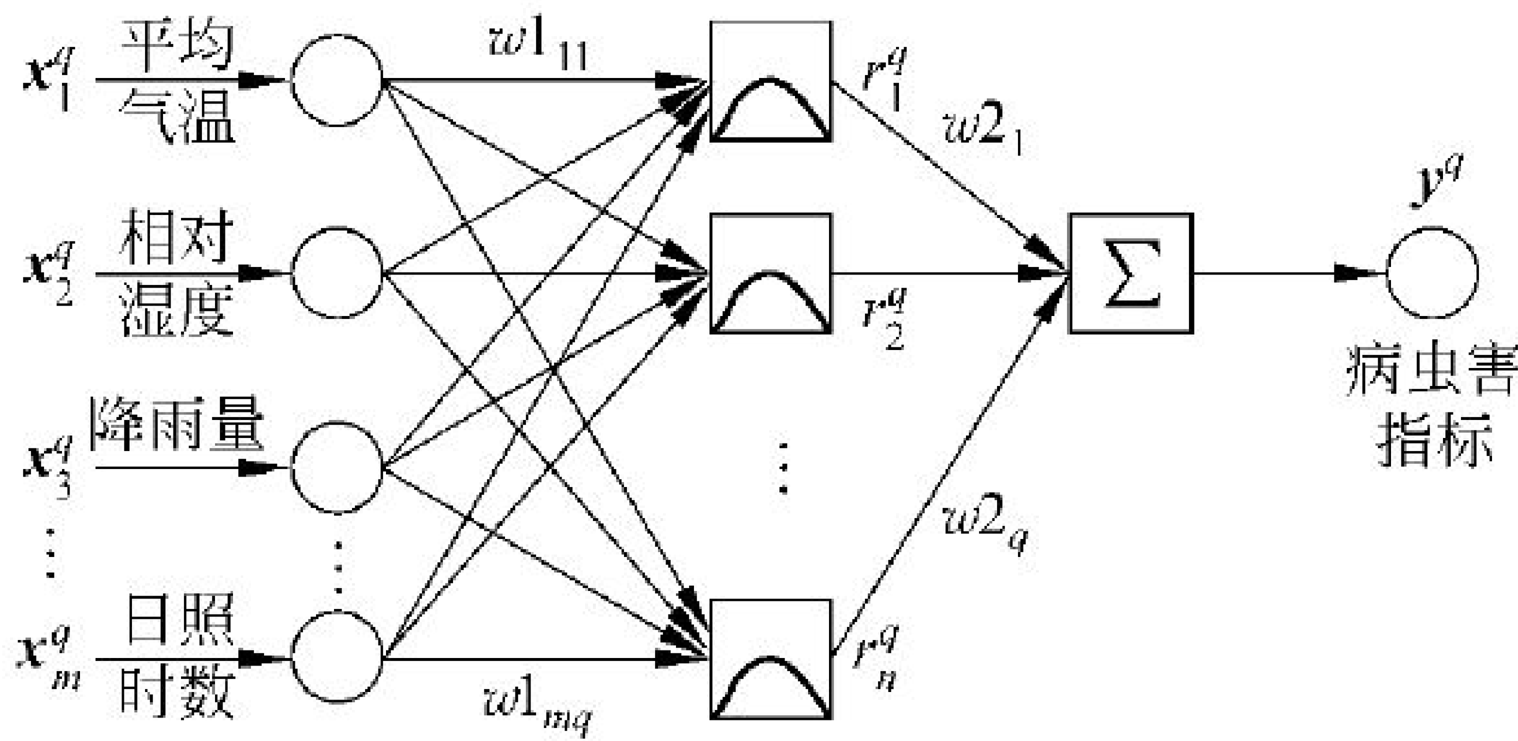


图 12-7　预测模型的径向基网络结构

径向基网络的训练过程实际上是确定隐层到输出层的连接权、隐层节点中心以及域场宽度、偏移量 w_0。径向基网络的训练算法有很多种，此处采用最广泛的正交最小二乘法观察网络仿真结果。

2. 梨黑星病病叶率径向基网络预测

用安徽省萧县 1999—2006 年梨黑星病的统计数据和气象资料作为数据集，以梨黑星病病叶率作为输出向量，以平均气温、相对湿度、日照时数、雨日、降雨量 5 个气象因子作为输入向量，建立网络预测模型。

梨黑星病病叶率数据共有 8 年 104 组，选 94 组数据作为训练集建立网络模型，选 10 组数据作为测试集。应用 Matlab 编写径向基网络预测程序，关键是确定合适的 spread 值，通过编程循环测试，取 spread＝1500，测试集原始数据及预测数据参见表 12-2。病叶率实测值与径向基网络预测值比较如图 12-8 所示。以误差≤5％为正确预测，病叶率径向基网络模型正确率为 80％。

表 12-2　梨黑星病病叶率测试数据集

测试样本	逐旬平均气温/℃	逐旬相对湿度/％	逐旬雨日	逐旬日照时数	逐旬降雨量	实测病叶率/％	RBF 预测值/％
1	24.9	54	0	76.3	0	3.7	4.01
2	28.2	58	2	87.1	19.5	9.3	3.18
3	25.8	82	6	35.5	227.3	8.9	8.12
4	26.3	87	4	26.8	224.6	5.8	7.65
5	28.7	83	3	43.5	6.8	13.8	8.25
6	26.7	84	4	41.3	35.7	18.6	17.38
7	27.4	84	3	50.6	8	16.1	15.20
8	28.0	77	2	84.3	9.7	15.1	16.06
9	26.5	86	7	48.2	49.2	12	13.23
10	24.9	54	0	76.3	0	12.9	17.05

3. 梨黑星病病果率径向基网络预测

用安徽省萧县 1999—2006 年梨黑星病的统计数据和气象资料作为数据集，梨黑星病病果率数据共有 8 年 74 组，选 64 组数据作为训练集建立网络模型，选 10 组数据作为测试集，以梨黑星病病果率作为输出向量，以病叶率、平均气温、相对湿度、日照时数、雨日、降雨量 6 个指标作为输入向量，建立网络预测模型。

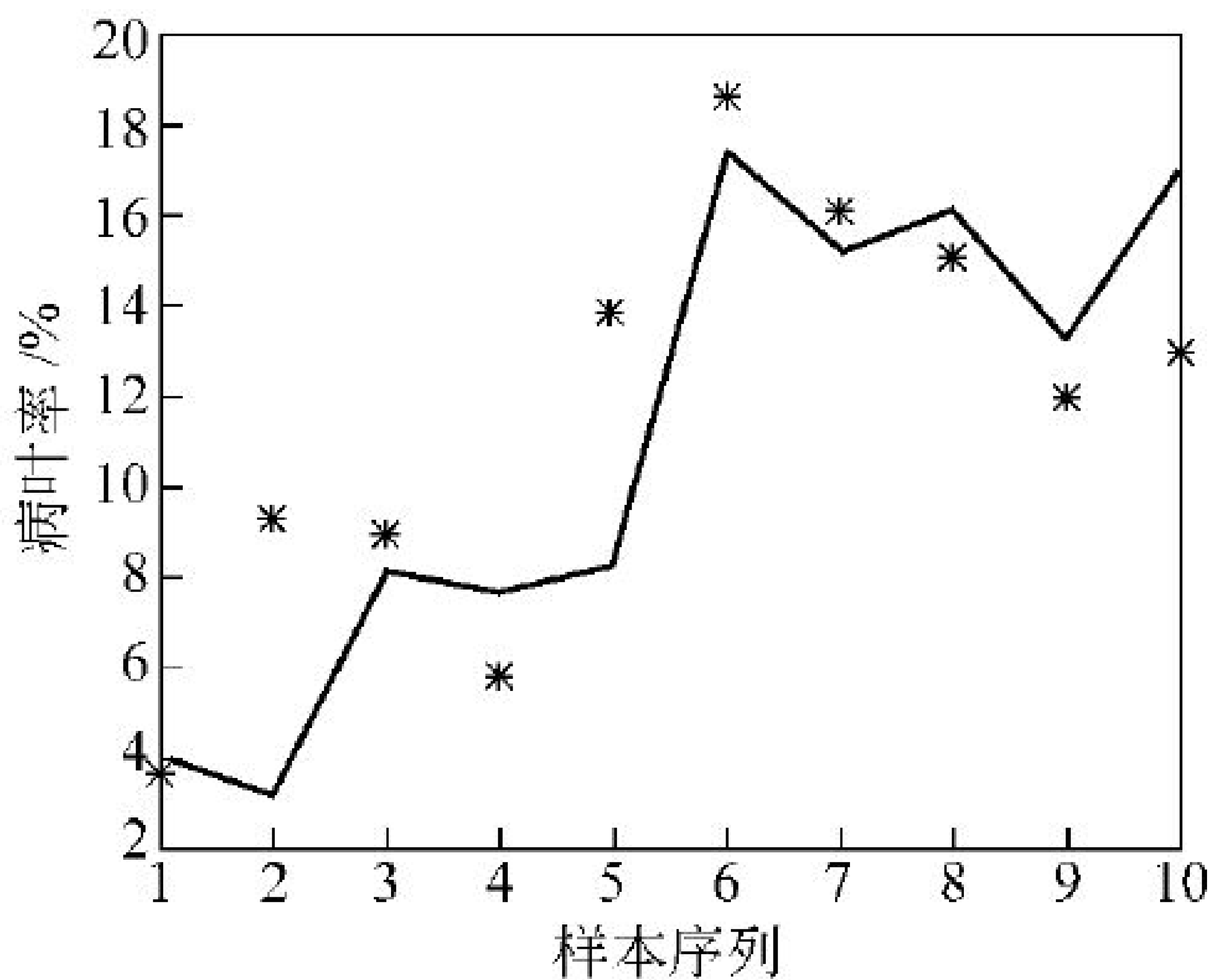


图 12-8　梨黑星病病叶率实测值与 RBF 网络预测值比较

应用 Matlab 编写径向基网络预测程序，通过编程循环测试，取 spread＝800，测试集原始数据及预测数据参见表 12-3。病果率实测值与径向基网络预测值比较如图 12-9 所示。以误差≤3％为正确预测，病果率径向基网络模型正确率为 80％。

表 12-3　梨黑星病病果率测试数据集

测试样本	逐旬平均气温/℃	逐旬相对湿度/％	逐旬雨日	逐旬日照时数	逐旬降雨量	病叶率/％	病果率实测值/％	RBF 网络预测值/％
1	25.4	69	2	67.9	55.7	10.8	0	0.13
2	25.4	72	2	54.8	71.5	6.8	0	2.50
3	25.1	78	3	94.6	48.5	16.5	0.3	2.60
4	26	77	2	11.3	61.1	11.4	0.8	4.97
5	29.2	76	0	0	89.5	9.8	1.6	4.76
6	27.4	78	2	34.9	79.7	7.3	2.1	4.94
7	26.3	82	4	8.4	53.6	9.3	4.3	6.79
8	23.3	82	3	54.5	54.6	10.3	6.4	6.40
9	24.3	72	2	43.5	84.3	1.5	0	2.97
10	26.5	54	1	0.4	104.8	5.2	0	2.40

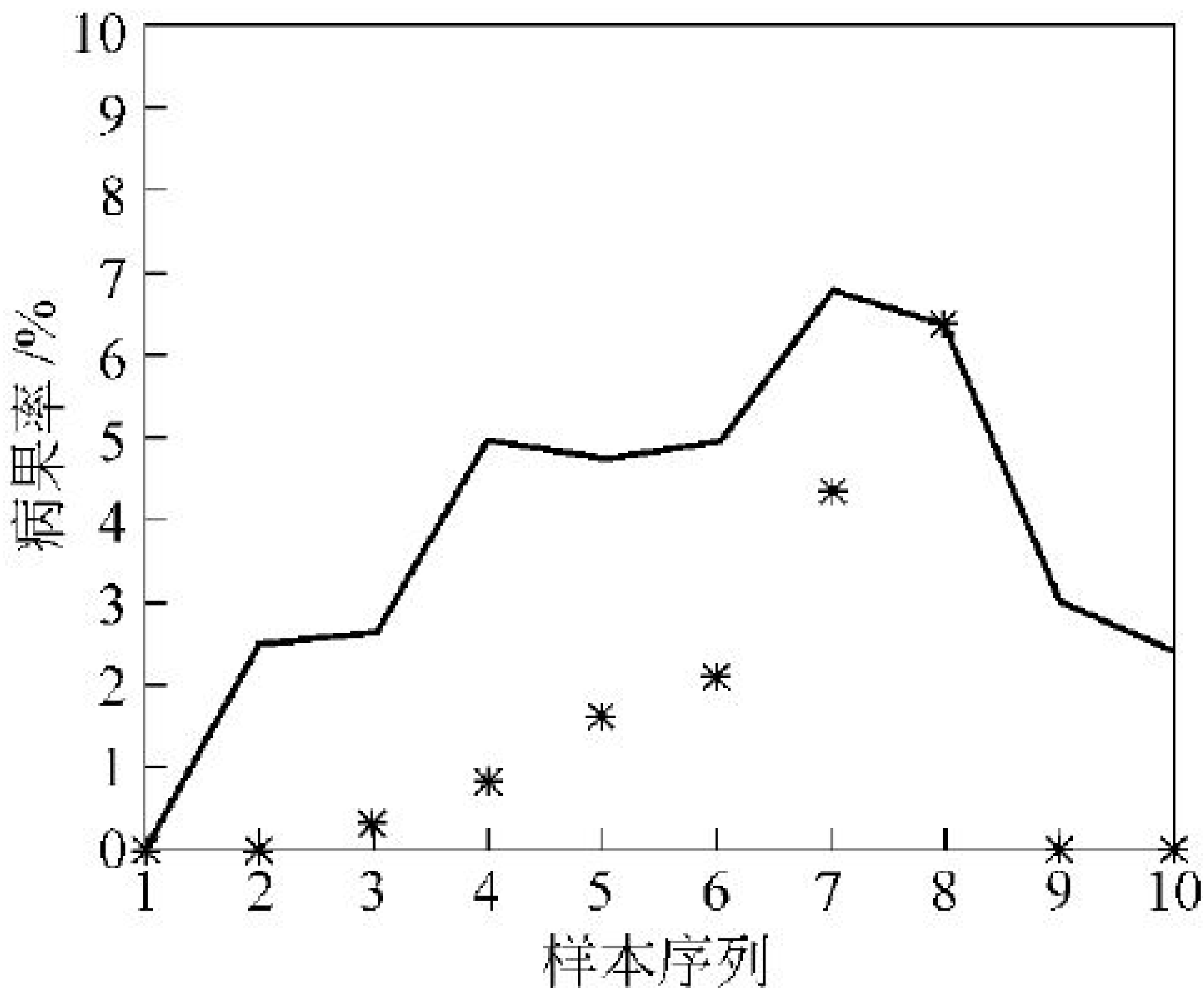


图 12-9　梨黑星病病果率实测值与 RBF 网络预测值比较

12.3.4　梨黑星病果实病指径向基网络预测

用安徽省萧县 1999—2006 年梨黑星病的统计数据和气象资料作为数据集，梨黑星病病果



率数据共有 8 年 74 组，选 64 组数据作为训练集建立网络模型，选 10 组数据作为测试集，以梨黑星病果实病指作为输出向量，以病叶率、病果率、平均气温、相对湿度、日照时数、雨日、降雨量 7 个指标作为输入向量，建立网络预测模型。

应用 Matlab 编写径向基网络预测程序，通过编程循环测试，取 spread＝1031.5，测试集原始数据及预测数据参见表 12-4。果实病指实测值与径向基网络预测值比较如图 12-10 所示。如果以误差≤1.5％为正确预测，果实病指径向基网络模型正确率为 100％；以误差≤1.0％为正确预测，果实病指径向基网络模型正确率为 80％。

表 12-4　梨黑星病果实病指测试数据集

测试样本	逐旬平均气温/℃	逐旬相对湿度/％	逐旬雨日	逐旬日照时数	逐旬降雨量	病叶率/％	病果率/％	果实病指/％	RBF 网络预测值/％
1	25.6	60.6	0	0	30.8	5.8	0.6	0.12	0.53
2	25.1	59.8	1	7.9	77.4	4.9	1.1	0.21	0.20
3	25.7	82	4	153	56	2.7	1.3	0.26	0.019
4	24.6	36	5	92.6	22.4	11.1	1.8	0.35	0.69
5	23.6	37	9	99.4	16.4	8	9	1.89	2.91
6	30.1	34.4	3	21.5	58.2	4.9	22	7.59	6.12
7	29.1	87.6	3	28.6	44.7	5.8	34.3	9.99	10.28
8	22.8	91.8	8	72.2	1.9	9.8	46.8	14.71	13.92
9	24.6	93	7	129.6	28.2	16.4	64.3	19.73	19.16
10	27.8	72.2	6	57.1	85.7	3.6	0	0	0

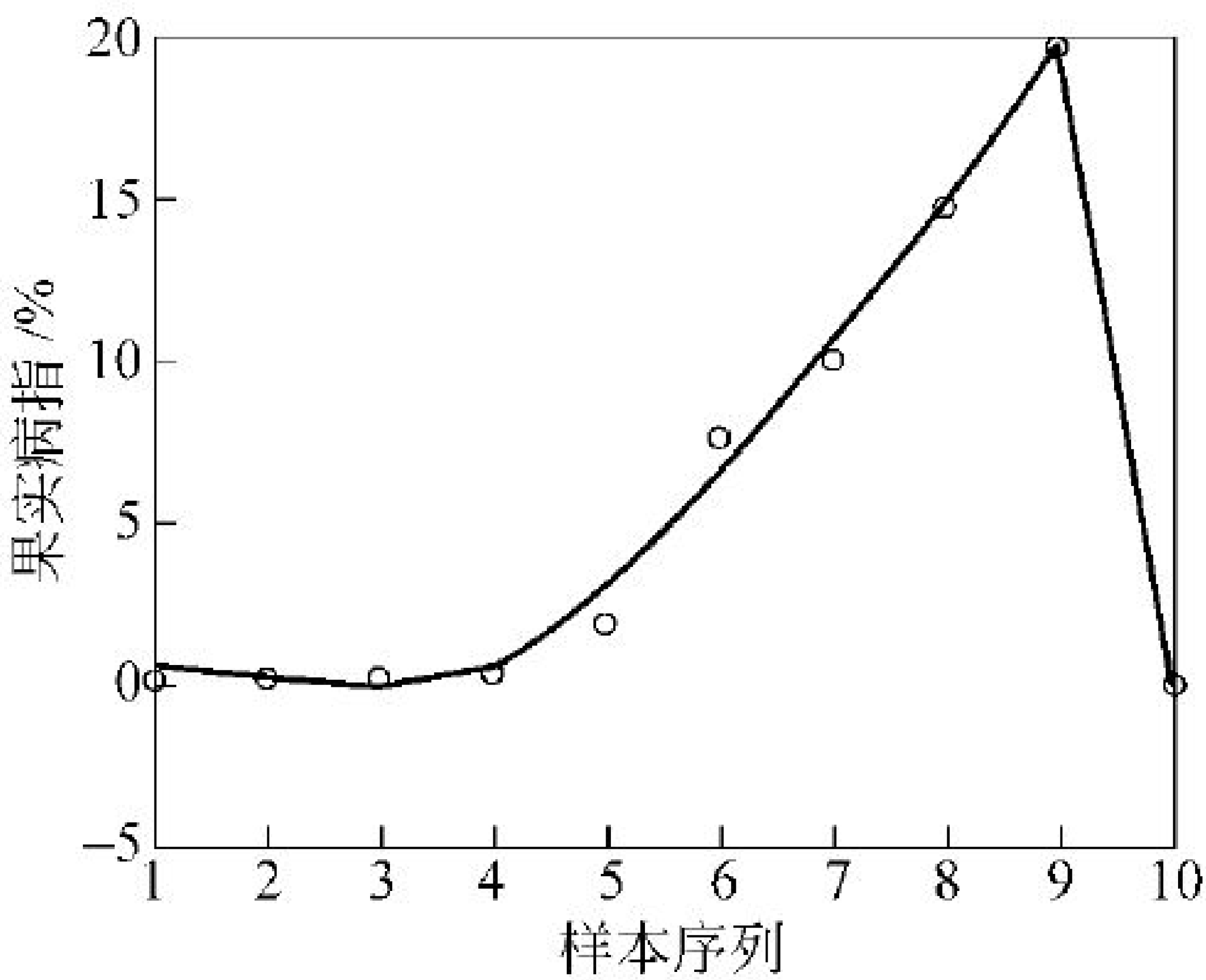


图 12-10　梨黑星病果实病指实测值与 RBF 网络预测值比较

12.3.5　梨黑星病的 BP 网络与径向基网络预测结果对比分析

（1）BP 网络预测

应用 Matlab 编写 BP 网络预测程序，其训练过程如图 12-11 所示。

（2）径向基网络预测分析

应用 Matlab 编写径向基网络预测程序，预测结果与 BP 神经网络预测值及黑星病的实际发生等级比较参见表 12-5，径向基网络的预测准确率为 80％，比 BP 神经网络的准确率高。

试验表明：神经网络方法良好的非线性逼近能力，可作为复杂多因子病虫的预测分析方法。

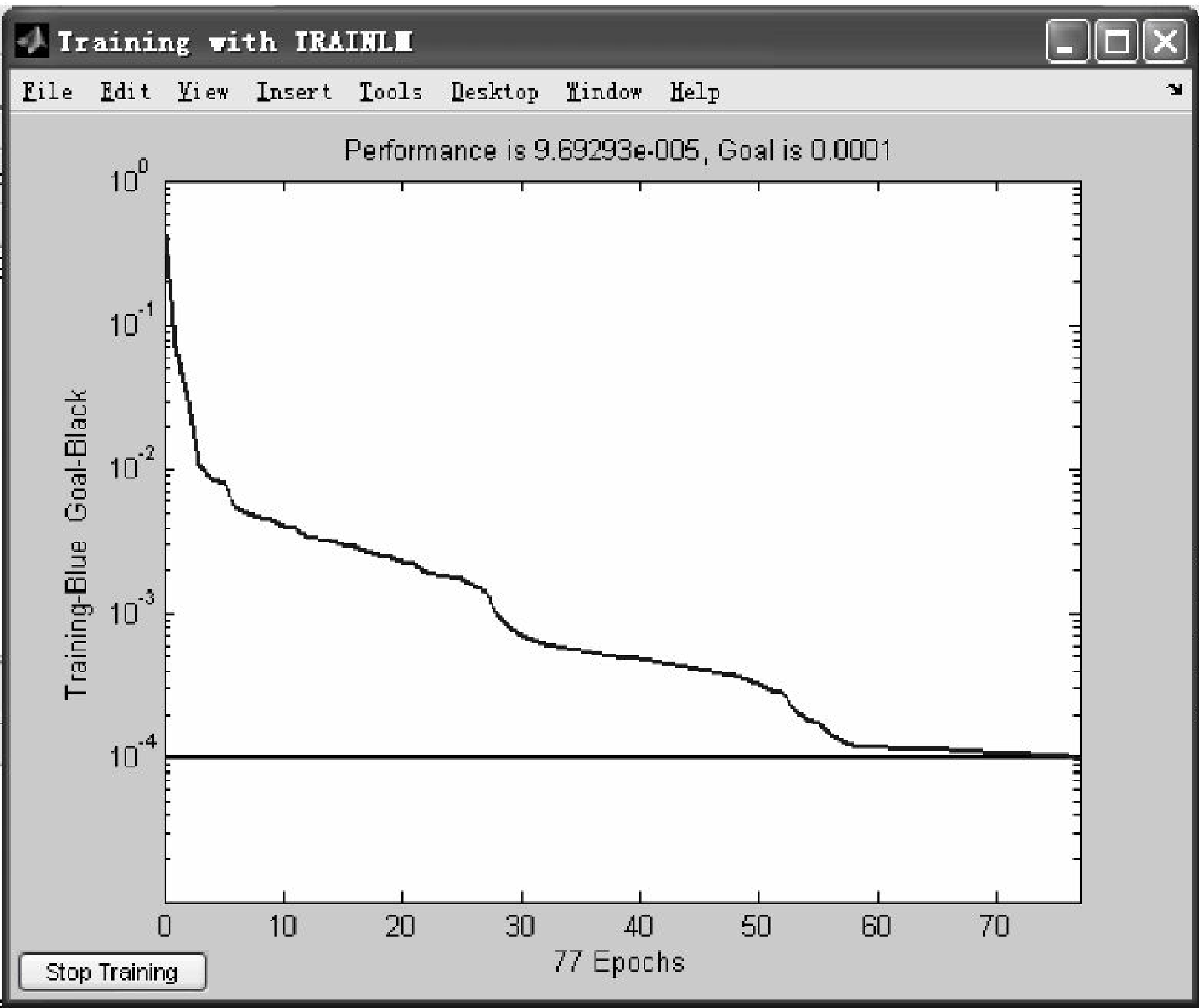


图 12-11　梨黑星病的 BP 网络训练过程

表 12-5　不同预测模型病叶率预测比较

预测与实测比较		病叶率/%										评价指标
		1	2	3	4	5	6	7	8	9	10	
实测值		3.7	9.3	8.9	5.8	13.8	18.6	16.1	15.1	12.0	12.9	
预测值	BP 网络	5.94	2.46	6.6	94.67	10.19	10.48	9.33	10.98	34.25	10.97	
	RBF 网络	4.01	3.8	8.12	7.65	8.25	17.38	15.20	16.06	13.23	17.05	
预测误差	BP 网络	2.24	6.84	2.3	88.87	3.61	8.12	6.77	4.02	22.25	1.93	大
	RBF 网络	1.31	5.5	0.78	1.85	5.55	1.22	0.9	0.96	1.23	4.15	小
预测结果	BP 网络	√	×	√	×	√	×	×	√	×	√	
	RBF 网络	√	×	√	√	×	√	√	√	√	√	80%

注：规定误差±5%以内为准确。

12.4　径向基神经网络泛化能力分析

泛化能力是指经训练后的网络对未在训练集中出现的样本(来自同一分布)给出正确反应的能力。学习不是单纯的记忆已学习过的输入，而是通过训练样本学习到隐含在样本中的规律性，从而对未出现的输入也能给出正确的反应。

泛化能力是神经网络最主要的性能，标志着神经网络的应用价值。如果把学习过程看作一个曲线拟合过程，泛化相当于非线性内插。给定的样本通常分成两组，一组用于训练网络，称为训练集；另一组用于检验学习结果，称为测试集。正确训练的网络，即使对训练样本有一点误差，但可对未出现的输入给出正确反应。相反，若学习了过多的特殊样本，且过分追求训练集内误差小，就会丧失泛化能力，出现所谓过拟合现象。这是因为学习过多时，往往记住了个别特例以至某些噪声，而未学到真正的规律。在检验网络性能时，应当用训练集以外的样



本,即测试集中不包括训练集中的样本。所以合理划分样本,使得既充分利用有限的样本去训练网络,又能较好地检验网络训练效果。

综上所述,影响泛化能力的因素有三个方面:训练样本的数量与质量、网络结构(规模)以及问题本身的复杂程度。

径向基神经网络利用自身网络结构(输入层节点、中间隐含层以及输出层节点)特点来进行数据预测和仿真,实际上是通过调节网络中心和输出层权值,使网络误差达到最小,其实质是一种非线性映射,它主要有以下特点:

(1) 网络构建容易。不需要了解因变量与自变量之间确切的数学表达关系,只要能确定出输入变量与输出变量,构建合理的网络模型,采用合理的算法,便可进行数据仿真和预测,因此更能适应较复杂的系统。

(2) 较高的收敛精度和收敛效率。与传统的数理统计回归分析和 BP 网络模型相比,具有较强的数值逼近能力,能快速收敛到全局最优,而不是局部极值,具有收敛速度快、建模效率高等特点。

(3) 网络较为敏感。当采用不同的隐含层神经元节点时,输出结果将会有较大误差,而且与输入自变量的选取有一定关系,一定程度上取决于设计者的实践经验。

利用径向基神经网络建立多因素时间序列预测模型,神经网络的输入节点个数由预测指标的影响因素个数确定,因此,从减少神经网络的输入节点个数出发的方法可以从根本上降低神经网络规模,提高网络泛化能力。

12.5　小结

本章通过分析梨黑星病病叶率、病果率和果实病指实测数据的梨黑星病时间动态,确定梨黑星病的中短期预测指标为病叶率、病果率和果实病指。在分析研究径向基网络算法和建模过程的基础上,认为梨黑星病发病指标与气象因子呈复杂非线性关系,适合采用具有非线性拟合能力的径向基神经网络,构建了基于径向基网络的梨黑星病病叶率、病果率和果实病指中短期预警模型,通过参数选择优化了算法。试验数据研究表明:径向基网络模型的预测结果优于 BP 神经网络的预测效果,病叶率、病果率、果实病指的预测准确率均达到 80%以上。

参考文献

1. Jiawei Han, Micheline Kamber. 数据挖掘概念与技术[M]. 北京,机械工业出版社,2006:327～340
2. 汪泗水,张孝羲. 基于神经网络的稻纵卷叶螟长期预测[J]. 植物保护学报,2000,27(4):313～316
3. 汪泗水,张孝羲,张夕林. 稻纵卷叶螟发生程度的神经网络预警[J]. 生物数学学报,2003,18(1):93～97
4. 李祚泳,彭荔红. 基于人工神经网络的农业病虫害预测模型及其效果检验[J]. 生态学报,1999,19(5):759～762
5. 王杰,檀根甲,胡易冰等. 基于神经网络的稻白叶枯病中期预警[J]. 安徽农业大学学报,2002,29(1):12～15
6. 高宁. 基于 BP 神经网络的农作物虫情预测预报及其 MATLAB 实现[D]. 安徽农业大学,2003

第 13 章

基于神经网络和遗传算法的牙鲆疾病预警模型

本章采用德尔菲法分析确定牙鲆疾病预警指标体系，采用神经网络和遗传算法相结合的方法，构建牙鲆疾病预警模型，包括牙鲆疾病的水环境预警、非水环境预警和症状预警。

13.1 牙鲆疾病预警指标体系

13.1.1 牙鲆疾病预警指标的确定

在水产领域专家的指导下，去除对牙鲆疾病影响小并且不易检测的预警指标，选取溶氧、水温、鱼体损伤、发病症状等 11 项预警指标作为评价牙鲆疾病预警警级的预警因子。对于这些因子，各个专家对其在牙鲆疾病预警中的重要程度认知并不一致，基于上述原因，我们采用指标重要程度的排序方法，通过专家问卷调查来确定最终对牙鲆预警进行总体评价的预警指标。在专家的指导下选择 11 个因素，请专家对这 11 个因素的重要程度进行排序。

13.1.2 牙鲆疾病预警指标重要程度排序

预警指标重要性程度系数的确定是对牙鲆疾病预警警级评价的关键环节之一。预警指标重要程度集 $\underset{\sim}{A}$ 确定得恰当与否，直接影响疾病预警的结果。预警指标重要程度的确定方法包括德尔菲法、专家调查法和判断矩阵分析法等，我们利用文献中介绍的方法，选用德尔菲(Delphi)法确定预警指标的重要程度，进而确定预警评价指标体系。预警指标重要程度系数的确定方法如下：

设牙鲆预警指标从下列 m 个指标中选取适当的因素进行评价，有溶解氧、水温、pH 值、浊度、饵料质量、鱼体损伤等因素，设预警指标集为

$$U = \{u_1, u_2, \cdots, u_m\} \tag{13-1}$$

U 的预警指标重要程度集为

$$\underset{\sim}{A} = \{a_1, a_2, \cdots, a_m\} \tag{13-2}$$

德尔菲法称为专家评议法，是利用多个专家提出的意见和建议来确定各个预警指标在疾病预警中的重要程度系数，用集合 $a_i(i=1,2,\cdots,m)$表示。求因素重要程度系数 a_i 的工作，必须由专家来进行，要求专家不但要有渊博的专业知识，而且要熟悉和掌握所研究问题的全部具体情况。

参加问卷调查的专家根据本人的经验和意见，确定各指标 u_i 的重要性序列值 $T_i(i=1,2,\cdots,m)$。T_i 值是 $1,2,\cdots,m$ 这 m 个数中间的某个数，即 $T_i\in\{1,2,\cdots,m\}$，并且要求每个专家不能给不同的预警指标打同样的分数(即 T 值)。最重要的指标的 T 值为 1，最不重要的指标



的 T 值为 m。将第 k 个专家就因素 u_i 所给定的因素重要性序列值记为 $T_{i\text{-}k}$。每一位专家提供一份各因素 u_i 的 T_i 值评定表。

根据集约化牙鲆养殖的具体条件，在选择评价指标时，把无关的和次要的影响因素去掉，保留如表 13-1 所列的 11 项预警指标，然后在这 11 项因素中确定最终的评价预警指标。参加评议的有 10 位专家，10 位专家所提供的 T_i 值的评定表可参见表 13-1。

表 13-1　专家对影响牙鲆发病因素重要程度评定表

专家号	因素 u_i										
	溶氧	pH 值	盐度	氨氮	水温	COD	鱼体损伤	饵料质量	发病症状	浊度	养殖密度
1	1	6	8	7	2	9	10	4	5	11	3
2	1	5	8	7	2	9	3	4	5	10	6
3	3	10	8	6	5	7	2	9	1	11	4
4	2	8	4	7	1	9	10	3	5	11	6
5	1	8	9	7	3	11	5	2	4	10	6
6	4	9	8	6	1	11	7	3	2	10	5
7	1	9	10	6	4	8	5	3	2	11	7
8	3	8	7	6	1	10	9	4	2	11	5
9	4	5	6	7	2	10	3	9	1	11	8
10	1	7	9	6	5	11	8	2	3	10	4

COD(chemical oxygen demand)：化学需氧量。

(1) 牙鲆预警因子优先得分表的获得

按专家 k 提供的因素重要性序列值 T_i 进行如下统计。

当 $\dfrac{T_{i\text{-}k}}{T_{j\text{-}k}}<1$ 时，记 $a_{ij\text{-}k}=1(i\neq j)$；

当 $\dfrac{T_{i\text{-}k}}{T_{j\text{-}k}}\geqslant 1$ 时，记 $a_{ij\text{-}k}=0(i\neq j)$；

假设参加评议的专家共有 n 位，将所有参加评议专家的 $A_{ij\text{-}k}$ 值进行累加，即

$$a_{ij}=\sum_{k=1}^{n}a_{ij\text{-}k},\quad i=1,2,\cdots,m;j=1,2,\cdots,m;i\neq j \tag{13-3}$$

这样得出 $m\times m$ 矩阵 $\mathbf{A}_{m\times m}$，如表 13-2 所示。

表 13-2　预警指标重要程度表

	溶氧	pH 值	盐度	氨氮	水温	COD	鱼体损伤	饵料质量	发病症状	浊度	养殖密度
溶氧		10	10	10	6	10	8	9	6	10	10
pH 值	0		6	3	0	8	4	1	1	10	2
盐度	0	4		2	0	8	3	2	2	10	2
氨氮	0	7	8		0	10	5	2	1	10	2
水温	4	10	10	10		10	10	7	6	10	8
COD	0	2	2	0	0		1	2	1	6	0
鱼体损伤	2	6	7	5	0	9		3	1	10	5
饵料质量	1	9	8	8	3	8	7		5	10	7
发病症状	4	9	8	9	4	9	9	5		10	8
浊度	0	0	0	0	0	4	0	0	0		0
养殖密度	0	8	8	8	2	10	5	7	2	10	

COD(chemical oxygen demand)：化学需氧量。

(2) 预警因子重要程度值

将上表中各行的 a_{ij} 值累加起来，得到

$$a_i = \sum_{j=1}^{m} a_{ij}, \quad i = 1,2,\cdots,m \tag{13-4}$$

a_i 表示第 i 行的 a_{ij} 的累加值，令

$$a_{\max} = \max(a_1, a_2, \cdots, a_m) \tag{13-5}$$

$$a_{\min} = \min(a_1, a_2, \cdots, a_m) \tag{13-6}$$

显然，与 $a_{\min}$ 相对应的因素的重要程度最低，而与 $a_{\max}$ 相对应的因素的重要程度同其他诸因素相比是最高的。

将表 13-2 中各行 a_{ij} 值的累加值 a_i($i=1,2,\cdots,11$)记录在表的第 12 列，如表 13-3 所示。可以看出，有 89 人次认为因素溶氧比其他因素重要。

表 13-3　预警指标重要程度排序结果表

	溶氧	pH 值	盐度	氨氮	水温	COD	鱼体损伤	饵料质量	发病症状	浊度	养殖密度	a_i	c_i
溶氧		10	10	10	6	10	8	9	6	10	10	89	1.1
pH 值	0		6	3	0	8	4	1	1	10	2	35	0.4647
盐度	0	4		2	0	8	3	2	2	10	2	33	0.4412
氨氮	0	7	8		0	10	5	2	1	10	2	45	0.5824
水温	4	10	10	10		10	10	7	6	10	8	81	1.0059
COD	0	2	2	0	0		1	2	1	6	0	14	0.2176
鱼体损伤	2	6	7	5	0	9		3	1	10	5	46	0.5941
饵料质量	1	9	8	8	3	8	7		5	10	7	65	0.8176
发病症状	4	9	8	9	4	9	9	5		10	8	71	0.8882
浊度	0	0	0	0	0	4	0	0	0		0	4	0.1
养殖密度	0	8	8	8	2	10	5	7	2	10		60	0.7588

COD(chemical oxygen demand)：化学需氧量。

从表 13-3 中得知，$a_{\max}=a_1=89$　$a_{\min}=a_{14}=4$，说明预警指标溶解氧的重要程度最高，而预警指标浊度的重要程度最低。

(3) 因子间重要程度级差 d

令 $b_{\max}=1$，$b_{\min}=0$，则

$$d = \frac{a_{\max} - a_{\min}}{b_{\max} - b_{\min}} \tag{13-7}$$

根据式(13-7)，得出

$$d = \frac{89-4}{1-0} = 85$$

(4) 预警因子重要程度系数 c_i

$$c_i = \frac{a_i - a_{\min}}{d} + 0.1, \quad i = 1,2,\cdots,m \tag{13-8}$$

由式(13-8)可得预警因子重要程度系数的集合

$$\underset{\sim}{C} = \{c_1, c_2, \cdots, c_m\} \tag{13-9}$$

按照式(13-9)，求出所要确定的集约化牙鲆养殖预警指标重要程度系数的集合为



$$\underset{\sim}{C} = (1.1, 0.4647, 0.4412, 0.5824, 1.0059, 0.2176, 0.5941, 0.8176, 0.8882, 0.1, 0.7588)$$

根据专家调查问卷的排序结果，结合实际情况和专家的建议与意见，选取系数大于 0.4 的 9 个因素作为预警指标，即溶氧、pH 值、盐度、氨氮、水温、鱼体损伤、饵料质量、发病症状、养殖密度，计算结果如表 13-3 所示。

13.1.3　牙鲆疾病预警指标警限确定

根据上节所述的指标选取，利用专家调查方法，确定集约化牙鲆养殖的主要影响因素为以溶氧、水温、盐度、氨氮、pH 值 5 个水环境参数为准的预测预警。对于牙鲆养殖来说，对水环境有特定的要求，为了满足预警评价的要求，对每一个影响因子，需要根据专家调查的方法，综合多个水产养殖专家的意见，来确定每个水质参数的无警、中警、重警的边界点，进而确定每一个警级的警级区间，警级边界点确定的公式为

$$V = \frac{\sum E(w_i) \times a_i}{N}, \quad i = 1, 2, \cdots, N \tag{13-10}$$

其中 $E(w_i)$为每个专家所占的权重大小，它的大小主要是根据专家的工作经验、年龄、学历等参数确定的，经与专家讨论后确定工作经验占 0.5、年龄占 0.2、学历占 0.3，a_i 为每个专家对警级边界点设定的值，N 为专家数量。由式(13-10)最终根据 4 位专家给出的信息得到每个水质参数的无警、中警、重警区间。饵料质量、鱼体损伤、水温溶氧等各因素警限值如表 13-4 所示。

表 13-4　预警指标警限表

指标名称	无　警	中　警	重　警
溶氧/(mg/L)	＞6	3～6	3 以下
水温/℃	12～23	24～26，8～11	＞26，＜8
COD	0～25	25～45	＞45
氨氮/(mg/L)	＜0.05	0.05～0.1	＞0.1
pH	7.25～8.25	6.25～7.25，8.25～8.62	＜6.25，＞8.62
盐度/%	15～32.5	32.5～40 或 10～15	＜8，＞40
鱼体损伤	低于 1.5%	1.5～5%	高于 5%
饵料质量	均衡营养、不变质，外形完好	营养缺失、受潮，部分破损、成份缺失、小部分霉变	霉变或虫蛀

13.2　牙鲆疾病的环境预警模型

牙鲆疾病的环境预警包括水环境状态预警、水环境趋势预警和非水环境预警三部分，其中牙鲆疾病的非水环境预警包括对饵料质量、鱼体损伤以及养殖密度的预警。牙鲆疾病水环境状态预警即对当前牙鲆所处的水质环境进行综合评价，确定牙鲆疾病的警级大小；趋势预警是根据现在的水质情况，对未来某一时刻水环境的各个预警因子进行预测，并利用状态预警的方法确定未来预警警级的大小。此外还将对各水质因子的变化趋势进行评价，若某个预警指标的监测值在一定时间内变化过快或过慢都表示水环境处于不正常的状态，下面分别对两种水

环境预警进行介绍。

13.2.1 牙鲆疾病水环境状态预警

水环境是牙鲆疾病预警的重要因素，是其赖以生存的条件，要使牙鲆正常代谢、正常生长，必须先要保持良好的水质环境。牙鲆疾病水环境预警指标分析可采用加权平均法、层次分析法和综合模糊评价方法等，但是加权叠加后有可能淡化水环境对疾病的影响。[1] 牙鲆疾病具有特殊性，即当一个指标不正常时就可能导致鱼的死亡。假设当某个预警指标十分差而其他指标均正常的情况下，这时可能评价结果是正常值，但由于计算方法忽略了单个指标的影响，因此可能导致对牙鲆疾病水环境警级的错误评价。

因此，本章对牙鲆疾病水环境状态预警的基本方法是根据各个水环境预警指标警限，将其进行排列组合，把水环境的各种可能情况都考虑进去，然后通过专家问卷调查的方法对每一种排列组合进行综合水环境警级的判定。最后综合所有专家给出的预警措施，制定相应预警预案。

牙鲆疾病的水环境因子预警的警度（WL）由溶氧（D）、水温（T）、盐度（S）、氨氮（N）、pH 值（P）共同决定。各个单因子指标的警级分为 3 级，状态预警的警级分为 5 级，警级的大小根据专家问卷调查的方法确定。D0、D1、D2 分别代表溶氧的等级为无警、中警、重警。WL 代表当前综合警级，0、1、2、3、4 分别为无警、轻警、中警、重警、巨警。根据 D，T，S，N，P 的不同组合，得出最终的预警警级。预警预案的制定则根据每个预警指标的单因子预警预案的组合确定。按照上述方法得出的牙鲆疾病水环境状态预警表如表 13-5 和表 13-6 所示，从表 13-6 可以看出处于无警和轻警的组合较少。

表 13-5 牙鲆疾病水环境状态预警表

D T S N P	WL	D T S N P	WL	D T S N P	WL	D T S N P	WL
D0T0S0N0P0	0	D0T0S0N1P0	1	D0T0S1N0P0	1	D0T0S2N2P0	3
D0T0S0N0P1	1	D0T1S0N1P1	3	D0T1S1N0P1	2	D0T1S2N2P1	2
D0T2S0N0P2	3	D0T2S0N1P2	4	D0T2S1N0P2	4	D0T2S2N2P2	3
D0T0S0N1P0	1	D0T0S1N2P0	1	D0T0S1N1P0	1	D0T0S2N0P0	3
D0T1S0N1P1	2	D0T1S1N2P1	3	D0T1S1N1P1	2	D0T1S2N0P1	4
D0T2S0N1P2	4	D0T2S1N2P2	3	D0T2S1N1P2	3	D0T2S2N0P2	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

表 13-6 牙鲆疾病水环境状态预警警级对照表

无警 0	轻警 1	中警 2	重警 3	巨警 4
D0T0S0N0P0	D0T0S1N0P0	D1T0S1N0P0	D0T0S2N0P0	D1T2S2N2P2
	D0T0S0N0P1	D0T0S1N0P0	D1T1S0N0P0	D2T1S1N2P1
	D0T0S0N1P0	D0T1S0N0P0	D1T0S2N0P0	D2T1S1N2P2
	D0T1S0N0P0	D0T1S0N1P0	D0T1S2N0P0	D2T2S0N1P2
		D1T0S0N1P0	D0T0S2N0P1	D2T2S1N2P2
		D0T1S0N0P1	D1T1S0N1P0	D2T2S2N1P2
		D0T1S1N0P0	D2T0S1N1P1	D2T2S2N2P2
		D0T0S1N1P0	D2T0S1N0P1	D2T2S0N2P0
		⋮	⋮	⋮



13.2.2 牙鲆疾病水环境趋势预警

水环境趋势预警是根据当前条件,预测未来水环境中各个因素的值,并对未来水质情况进行预警。此处拟采用BP神经网络与遗传算法相结合的方法对水质环境进行预测预警。下面先对此预警模型进行简要介绍。

人工神经网络(ANN)是一种模拟人的神经系统而建立起来的非线性动力学模型,由大量称为神经元的简单信息单元组成,每个神经元不仅从邻近的其他神经元接受信息,也向邻近神经元发出信息。整个网络的信息处理是通过神经元之间的相互作用来完成的。神经网络具有自组织、自学习和联想记忆功能,并具有分布式、并行性及高度鲁棒性等特点。[2]

BP(back propagation)即误差反传神经网络,是人工神经网络模型中预测效果较好的算法之一,但是BP算法基于梯度方法,最大的缺陷是收敛速度慢,对于较大的搜索空间,多峰值极易陷入局部极值点,从而影响预测精度。另外神经网络的初始连接权值、阈值以及网络结构的选择缺乏依据,具有很大的随机性,很难选取具有全局性的初始点,因而求得全局最优的可能性较小。[2]

遗传算法利用达尔文生物进化过程物竞天择、适者生存的思想,是一种对全局优化搜索的算法。它适用于多维复杂的非线性系统,具有适应性、全局优化性和隐含并行性等特点,有很强的全局搜索能力。遗传算法起源于20世纪60年代对自然和人工自适应系统的研究。最早由美国密执安大学的Holland教授提出。20世纪80年代由Gildberg进行了归纳总结,形成了遗传算法的基本框架。[3]

由于遗传算法是一种全局优化搜索算法,因此可以弥补BP神经网络易陷入局部极值点的不足。但是遗传算法也有不足之处,即当遗传算法搜索到最优解附近时,局部搜索能力不强,无法精确地确定最优解的位置,因此它对局部搜索空间的微调能力较差,而BP算法用于局部搜索情况则比较有效。因此,基于BP神经网络和遗传算法各自的优缺点,可以将两者结合起来进而充分发挥BP神经网络和遗传算法(GA)各自的长处,弥补相互之间的不足,从而达到较好的预测效果。[4,5]

1. BP神经网络基本原理

BP算法是一种基于梯度下降的最优化方法,属于有导师的学习方法。BP网络的学习过程分为两个阶段,包括正向的传播和反向传播两部分:第一阶段进行正向传播,每一层神经元的状态只影响下一层神经元结构,如果输出层不能得到期望输出,即期望输出与实际输出之间存在误差时,就进入第二阶段,进行反向传播。

BP神经网络将误差信号沿着原来的路径返回,并根据误差梯度下降原则来调节各层的权值和阈值,通过不断修改各层神经元的权值,使修改后的输出值能够接近期望值。修改以后的权值再经过正向传播过程,再将期望输出与实际输出进行比较。

按照上述方法不断循环往复直到输出值与期望值的误差达到可接受范围或者训练次数超过预设值,训练过程才结束。BP神经网络的神经元结构和网络结构如图13-1所示。

文献研究表明包含两个隐层的BP神经网络能以任意精度逼近任意连续函数及其各阶导数,[6]且具有多隐含层神经网络能够更好地逼近非线性函数,但是计算量会随之增加,并且有可能不稳定,而对于BP网络来说,若隐含层节点个数设置合理,单隐含层也能达到同样的效果。为了减少模型计算的复杂度,这里采用3层BP神经网络模型。

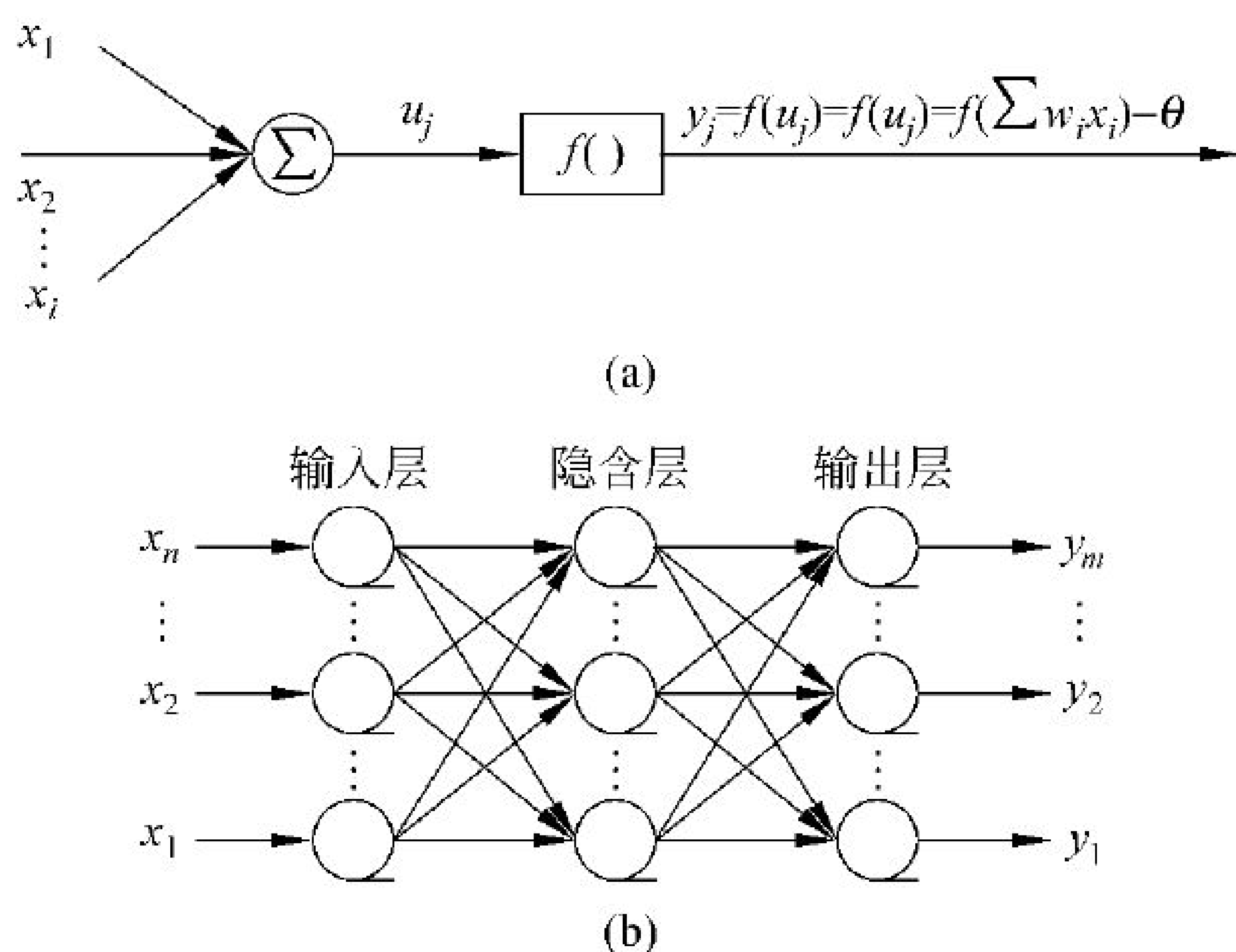


图 13-1　BP 神经网络结构图

下面将给出 BP 神经网络的计算过程。[7] 设输入层中有 n 个神经元，隐含层节点有 h 个神经元，输出层有 m 个神经元，样本输入值分别为 $(x_1, x_2, \cdots, x_n)$，样本输出值分别为 $(S_1, S_2, \cdots, S_m)$ 隐含层的输出值分别为 $(O_1, O_2, \cdots, O_h)$，BP 网络输出值分别为 $(y_1, y_2, \cdots, y_m)$，$w_{ij}(i=1,2,\cdots,n; j=1,2,\cdots,s)$ 为输入层与隐含层节点之间的权值，$v_{jk}(j=1,2,\cdots,h; k=1,2,\cdots,m)$ 为隐含层与输出层节点之间的权值，隐含层和输出层神经元的阈值为 θ_j 和 ϕ_k。

选择 Sigmoid 函数作为激活函数，计算隐含层节点的输出值：

$$H_j = \sum_{i=1}^{n} w_{ij} x_i + \theta_j, \quad j = 1, 2, \cdots, h \tag{13-11}$$

$$O_j = f(H_j) = \frac{1}{1 + \mathrm{e}^{-H_j}} \tag{13-12}$$

同理将隐含层节点的输出值作为输出层节点的输入值，并由式(13-13)和式(13-14)求出 BP 神经网络的输出值：

$$I_k = \sum_{j=1}^{h} f(H_j) v_{jk} + \phi_k, \quad k = 1, 2, \cdots, m \tag{13-13}$$

$$y_k = f(I_k) = \frac{1}{1 + \mathrm{e}^{-I_k}} \tag{13-14}$$

逐一地根据样本集中的样本 $(x_1, x_2, \cdots, x_n)$，$(S_1, S_2, \cdots, S_m)$ 计算出实际输出 y_k 和误差测度 E，根据式(13-15)和式(13-16)分别求出网络关于第 p 个样本的误差测度和网络关于整个样本集的误差测度：

$$E_p = \frac{1}{2} \sum_{k=1}^{m} (y_{pk} - S_{pk})^2 \tag{13-15}$$

$$E = \sum_{p=1}^{m} E_p \tag{13-16}$$

用输出层的误差调整输出层权矩阵，并用此误差估计隐含层的误差，再用隐含层误差估计输入层的误差。如此获得所有其他各层的误差估计，并用这些估计实现对权矩阵的修改。形成将输出端表现出的误差沿着与输入信号相反的方向逐级向输入端传递的过程。根据 δ 学习规则对各层间的权值进行调整，式(13-17)和式(13-18)为权值调整公式。公式中 η 为学习因子，t 为训练次数 w_{kj} 和 v_{kj} 代表本次训练结束后被修改的各层节点间的权值，δ 为输入、输出间



的误差变化。

$$w_{kj}(t+1) = w_{kj}(t) + \eta\delta_{pk}O_{pj} \tag{13-17}$$

$$v_{kj}(t+1) = v_{ji}(t) + \eta\delta_{pj}x_{pj} \tag{13-18}$$

判断样本输出与网络输出之前的误差，如果小于 ε(ε<0.001)则 BP 神经网络训练过程结束，否则继续进行权值和阈值的调整，再次计算误差大小，直至误差小于 ε 训练结束。

由于需要确定神经网络隐含层节点个数，并且避免使神经网络陷入局部极小值，可以采用遗传算法进行优化处理，具体处理方法将在下一节进行讨论。

2. 基于遗传-BP 神经网络的水环境趋势预警

遗传算法是一种基于概率的搜索算法，其基本思想是模拟由二进制和实数串组成的群体进化过程。具体操作是从由 n 个串组成的一个初始群出发，对当前群体中的个体循环地执行选择、交叉和变异过程，不断进行直至得到最优解。遗传算法程序运行流程如图 13-2 所示。

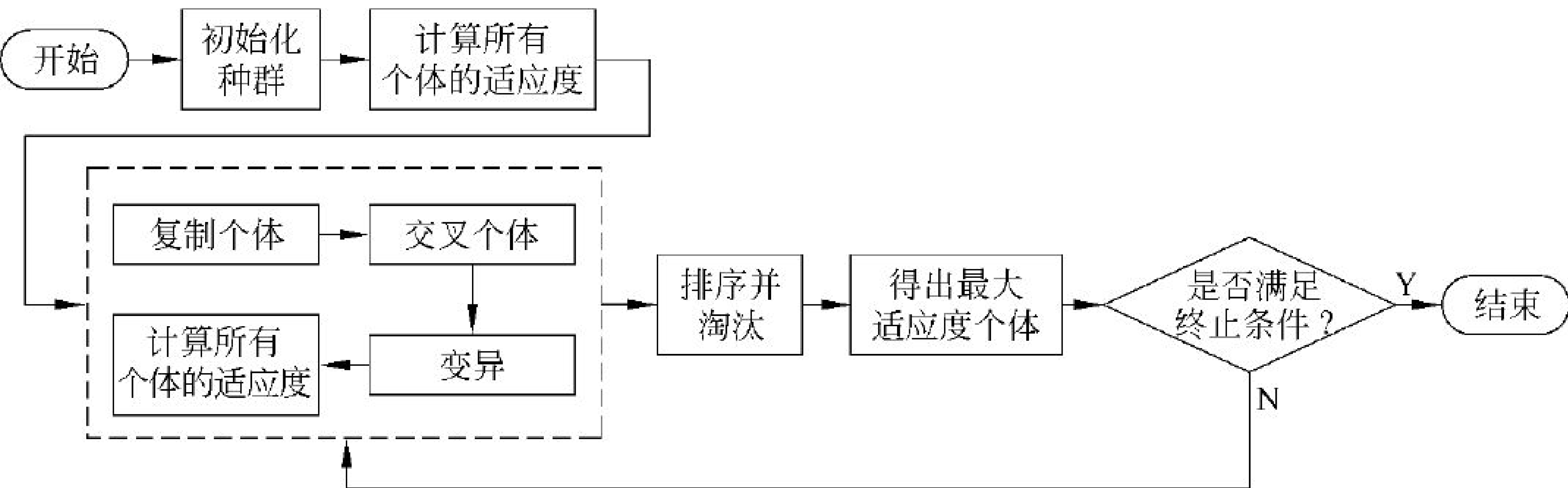


图 13-2　遗传算法流程图

为了提高神经网络的训练精度，用增加神经元数的方法要比增加隐含层的方法在结构实现上简单得多。因此网络设计是在尽可能少的层数前提下选取最佳神经元数，选取多少个隐含层节点才合适，在理论上并没有一个明确的规定。隐含层神经元数太少，容错性就会变差，因而不能识别训练样本集中不包含的样本。隐含层神经元数过多不但会使参数增多而导致学习时间过长，而且会导致对新样本的学习能力降低，使网络总体性能变坏。因此确定 BP 网络隐含层节点非常重要。

遗传-BP 神经网络主要是利用遗传算法全局搜索能力强的特点，先用遗传算法对网络的隐含层节点数量、权值和阈值进行全局粗精度的预学习，定位最优解区域并确定最优隐层个数，使得权值和阈值种群聚集在参数解空间的某几处，再用 BP 神经网络算法在这些小的解空间中进行梯度细搜索，最终求得最优解。遗传算法对 BP 神经网络的优化过程包括编码、解码、确定适应度函数、遗传操作（自然选择、交叉、变异等），下面具体介绍这个过程。

(1) 对 BP 神经网络中的权值和阈值采用实数编码，对隐含层节点数用二进行表示，0 代表该隐含层节点无效，1 代表有效。其中隐含层节点数的最大值，即隐含层染色体的个数可根据 Kolmogorov 定理，采用经验公式 $L=\text{int}[(n+o)^{1/2}]+a$ 求出，其中 L 为隐含层节点个数，n 为输入层节点个数，o 为输出层节点个数，a 为 1 至 10 范围内的整数。

(2) 将父代的种群进行复制，这样就产生一部分与父代完全相同的子代。然后对被复制

的父代进行交叉和变异操作。交叉在进化过程中起着十分重要的作用，是进行优化搜索必不可少的步骤，而为了避免种群的过早收敛，变异操作也是十分必要的。进行交叉和变异操作时，需对实数编码和二进制编码分别考虑。对于二进制编码来说，在进行交叉操作时只需产生一个随机交叉点，然后以交叉点为界交换染色体段。同理，在进行变异操作时随机产生一个变异点，以一定的变异率对此变异点进行变异，即 0 变 1，1 变 0 的操作。

对于实数编码来说，采用的交叉方法基于凸集理论，交叉操作方式如式(13-19)，变异操作方式如式(13-20)。

$$\left.\begin{array}{l} y_i^{(1)} = a \times x_i^{(1)} + (1-a) \times x_i^{(2)} \\ y_i^{(2)} = a \times x_i^{(2)} + (1-a) \times x_i^{(1)} \end{array}\right\} \tag{13-19}$$

$$\left.\begin{array}{ll} y_i = y_i + \beta \times (a_i - y_i), & P_m \geqslant \beta \geqslant 0.5 \\ y_i = y_i + \beta \times (b_i - y_i), & \beta < 0.5, \beta \leqslant P_m \end{array}\right\} \tag{13-20}$$

式(13-19)中，$x_i^{(1)}$，$x_i^{(2)}$，$y_i^{(1)}$，$y_i^{(2)}$ 分别为两个相邻父代和相邻子代的染色体的第 i 个变量，a 为每一对染色体的交叉率，为 0 至 1 范围内的随机数。式(13-20)中 y_i 为变异染色体的第 i 个变量；β 为 0 到 1 范围内的随机数，它随机地决定变异方向，P_m 为第 i 个变量的变异率，a_i，b_i 分别为权值、阈值的上下限，取值为＋1 和－1。

(3) 通过计算种群中所有染色体的适应度，从种群中选择总数的 1/2 的染色体数进化到下一代，由于父代在交叉变异之前进行过复制，所以淘汰 50%之后种群的数量不会随之改变。

通过淘汰优化后得到新的一代种群，进而返回至第二步，重新再进行复制、交叉、变异、选择操作。这样循环往复直到两代最优适应度函数相差很小或者达到预先设定的进化代数时，程序结束。其中适应度函数由式(13-21)给出。

$$F = a \times f_1 + b \times f_2 \tag{13-21}$$

其中 $0<a,b<1$ 且 $a+b=1$，我们选取 $a=0.8$，$b=0.2$，$f_1=1/(E+1)$，其中 E 为 BP 神经网络的误差，由式(3-6)得出，$f_2=L/m$，L 隐含层节点的最大值，m 为当前隐含层节点的个数。

对 BP 神经网络进行权值、阈值和隐含层节点数的优化后，就可以用 BP 神经网络训练样本，并最终确定预测模型，对未来水环境进行预测。遗传-BP 神经网络流程如图 13-3 所示。

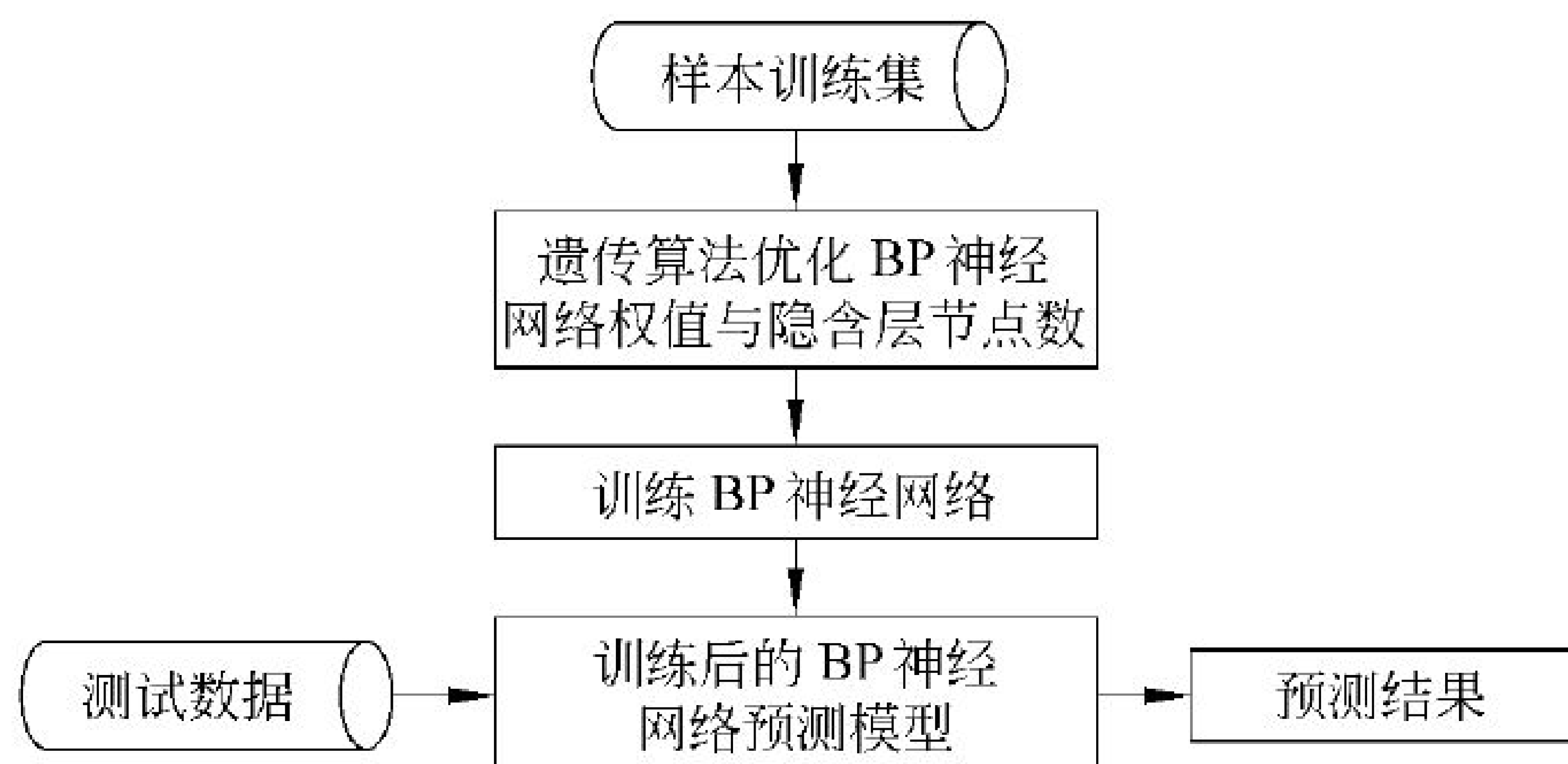


图 13-3　遗传-BP 神经网络流程

可将牙鲆疾病的水环境预警系统视为构建一个包括输入层、隐含层和输出层的三层 BP 神经网络。就网络结构而言，输入层神经元的个数由输入指标决定，能够对输出产生影响的输入指标包括水温、养殖密度、溶氧、氨氮，气压、pH 值、COD、盐度等。但是由于养殖密度在集



约化设施养殖场中是相对固定的，因此可去掉这个因子。由于实验条件所限，还不能达到对氨氮、COD的快速监测，而当前的溶氧和水温值可间接反映当前的氨氮和COD的大小，因此可忽略输入因子氨氮和COD。输出值为两小时后的水温、溶氧、pH值、盐度值。原始数据的矩阵如下，其中 $\boldsymbol{X}$ 为输入样本矩阵，$\boldsymbol{Y}$ 为输出样本矩阵。

$$\boldsymbol{X}=\begin{bmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{bmatrix}=\begin{bmatrix} x_{11},x_{12},x_{13},x_{14},x_{15},x_{16} \\ x_{21},x_{22},x_{23},x_{24},x_{25},x_{26} \\ \vdots \\ x_{n1},x_{n2},x_{n3},x_{n4},x_{n5},x_{n6} \end{bmatrix}$$

$$\boldsymbol{Y}=\begin{bmatrix} y_1^{\mathrm{T}} \\ y_2^{\mathrm{T}} \\ \vdots \\ y_n^{\mathrm{T}} \end{bmatrix}=\begin{bmatrix} y_{11},y_{12},y_{13},y_{14} \\ y_{21},y_{22},y_{23},y_{24} \\ \vdots \\ y_{n1},y_{n2},y_{n3},y_{n4} \end{bmatrix}$$

首先根据公式(13-22)对样本的输入和输出值进行归一化处理，使其值位于区间[0,1]。式(13-22)中的 $x_{\max j}$、$y_{\max k}$、$x_{\min j}$、$y_{\min k}$ 分别为样本中每个因子的最大值和最小值，x_{in}、y_{out} 为样本归一化处理后的输入和输出值。

令

$$\left.\begin{aligned} &\left.\begin{aligned} x_{\min j} &= \min(x_{1j},x_{2j},\cdots,x_{nj}), \\ x_{\max j} &= \max(x_{1j},x_{2j},\cdots,x_{nj}), \end{aligned}\right\} \quad j=1,2,\cdots,6; \\ &\left.\begin{aligned} y_{\min k} &= \min(y_{1k},y_{2k},\cdots,y_{nk}), \\ y_{\max k} &= \max(y_{1k},y_{2k},\cdots,y_{nk}), \end{aligned}\right\} \quad k=1,2,\cdots,4; \\ &\left.\begin{aligned} x_{\text{in}ij} &= \frac{x_{ij}-x_{\min j}}{x_{\max j}-x_{\min j}}, \\ y_{\text{out}ik} &= \frac{y_{ik}-y_{\min k}}{y_{\max k}-y_{\min k}}, \end{aligned}\right\} \quad i=1,2,\cdots,n \end{aligned}\right\} \tag{13-22}$$

处理后的样本数据按照前述方法，经过对模型的优化处理和对BP神经网络的训练，可以确定最终BP神经网络模型。

根据前述方法，用此遗传-BP神经网络对牙鲆疾病水环境因子进行训练，用200个样本集作为训练样本，30个样本集作为测试样本。通过不断调整权值和阈值，使 ε 值满足预测精度的要求。我们选用养殖场数据进行训练，首先利用遗传算法经过27代的进化后(其中初始化种群大小为100)，适应度值基本稳定，得到初始化的BP神经网络的权值和阈值，并得出隐含层节点数为5。然后对神经网络进行训练，经过2000次训练，样本与网络输出值间的误差已小于0.0028，训练成功，从而确定了训练成功后的BP神经网络模型。将训练好的神经网络用实际测试样本数据进行验证。由于输入值为0～1之间的数，经过BP神经网络模型计算后，得出的输出结果也是0～1之间的数，因此需要通过公式(13-23)对输出值进行反归一化，这样才能够得到最后的真实结果。

$$\left.\begin{aligned} x_{ij} &= x_{\text{in}ij}\times(x_{\max j}-x_{\min j})+x_{\min j} \\ y_{ik} &= y_{\text{out}ik}\times(y_{\max k}-y_{\min k})+y_{\min k} \end{aligned}\right\} \tag{13-23}$$

部分样本集如表13-7所示。

表 13-7　训练样本集

输　　入							输　　出				
时间	溶氧	水温	pH	室温	气压	盐度	时间	溶氧	水温	pH	盐度
11:40	6.1	24	7.85	27.6	992	28.32	13:40	5.72	24.2	7.84	28.2
11:45	6.1	24	7.85	27.6	992	28.32	13:45	5.73	24.2	7.84	28.2
11:50	6.09	24.1	7.85	27.6	992	28.32	13:50	5.74	24.2	7.84	28.26
11:55	6.06	24.1	7.84	27.6	992	28.2	13:55	5.74	24.2	7.84	28.26
12:00	6.05	24.1	7.84	27.6	992	28.2	14:00	5.75	24.2	7.85	28.26
12:05	6.03	24.1	7.84	28.1	992	28.2	14:05	5.76	24.3	7.85	28.26
12:10	6.01	24.1	7.84	28.1	992	28.2	14:10	5.77	24.3	7.86	28.2
12:15	5.98	24.1	7.84	28.1	992	28.2	14:15	5.77	24.4	7.86	28.2
12:20	5.96	24.1	7.84	28.1	992	28.2	14:20	5.77	24.4	7.86	28.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

13.3　基于症状的牙鲆疾病预警

本节将利用基于案例推理诊断与专家知识相结合的方法，对牙鲆疾病进行症状预警。症状预警包括疾病诊断和疾病预警两部分。先根据专家知识确定不同疾病在不同发病率和死亡率的警级大小。当用户输入症状时，首先根据用户所输症状对牙鲆疾病进行诊断，得出疾病诊断的结果；然后再结合用户所输的有此症状的牙鲆的发病率来确定症状预警警级。其中采用基于案例的方法对疾病进行诊断。

13.3.1　基于案例检索的牙鲆疾病诊断

目前，国内外对牙鲆疾病的研究主要是从牙鲆疾病的发病机理、发病症状和治疗方法等方面展开。但是对于牙鲆疾病诊断专家系统的研究成果还不是很多。基于规则的诊断推理方法和基于案例的诊断推理方法是在疾病诊断专家系统中应用比较广泛的方法。基于规则推理方法是根据以往专家诊断的经验，将其归纳成规则，通过启发式经验知识进行推理。只有在建立了较为完备的领域知识规则库的基础上，它才能具备较高的诊断正确率。[9]但在牙鲆疾病领域中，由于从领域专家中获取诊断规则比较困难，很难保证良好的诊断问题求解性能。

基于案例的推理是有别于基于规则推理的另一种推理模式，它是基于过去求解类似问题的经验来获得当前问题求解结果的一种新型专家系统推理模式。[10]对于牙鲆疾病诊断来说，利用水产专家诊断的历史案例来解决当前的新病例，案例可以从专家以往的历史诊断病例中获取。

（1）案例的存储

在进行牙鲆疾病案例诊断之前，首先应将案例进行有效的知识表示，即分析案例库的存储方式。传统的案例库表示是先将疾病症状进行编号，然后将每个症状作为案例库的一列，当案例中有此症状时值为 1 反之为 0。具体表现形式如表 13-8 所示。

上述案例表示方法需要在每条记录中存储每一个症状，这样会导致很多占用存储空间的空记录，因此增加了案例库的存储空间、降低了存取效率。而且当增加或减少疾病症状时，也必须对整个数据表进行重新设计，从而增加了系统维护的开销，给案例的管理和维护带来了不便。



表 13-8 症状-案例存储表

Caseid	Disease name	Problem description													cure
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	
Case1	Dis 1	1	1	0	0	0	1	1	1	0	0	0	0	0	Cure1
Case2	Dis 2	1	1	0	0	0	1	1	0	1	0	0	0	0	Cure2
Case3	Dis 3	0	0	1	0	0	0	0	0	0	1	1	0	0	Cure3
Case4	Dis 3	0	0	1	0	0	1	0	0	1	0	0	1	0	Cure4
Case5	Dis 2	0	1	0	1	0	0	0	0	0	0	0	1	1	Cure5
Case6	Dis 2	0	0	1	1	0	0	1	0	1	0	0	0	1	Cure6

为此,我们采用了另外一种存储模式(如表 13-9 所示),即将案例症状用数字或字母的编号表示,当进行案例存储时将案例集作为案例库中的一列,症状表示为案例中出现症状的数值组合。表中数值与症状名称的对应关系则存储在另外一个表中。增加或减少案例症状时,可以在症状表中增加一条数值与症状对应的记录,并在案例库中修改出现此症状的案例记录。进行案例检索匹配时,只需在程序中对 Case_sets 字段中的案例症状集作字符串拆分处理即可。这样的存储方法可以简化案例的存储空间,并能够增加对系统维护的效率。

表 13-9 案例集-案例存储表

Caseid	Disease name	Case type	Case sets	Cure	Prevent
Case1	Dis 1	1	1-4-15-23-31-32	Cure1	Prevent1
Case2	Dis 2	2	1-2-5-11-15-21-25-36	Cure2	Prevent2
Case3	Dis 3	3	1-3-4-7-16-32	Cure3	Prevent3
Case4	Dis 3	4	2-4-7-9-12-14-13-18-35	Cure4	Prevent4
Case5	Dis 2	5	3-5-9-10-13-15-13-20-24-29	Cure5	Prevent5
Case6	Dis 2	5	2-8-11-12-14-19-22-35	Cure6	Prevent6

(2) 案例的分类

基于案例的牙鲜疾病诊断主要分为两个步骤:案例分类和案例诊断的相似度的计算。其中案例分类主要是将相似度十分接近的案例划归为一类,从而将一个大案例库划分为若干个子案例库。这样做的目的是将相似的案例划分为一类,使得在基于案例的诊断过程中可以减少不相似案例给诊断带来的干扰,增加案例诊断的准确率。案例的分类算法为:先计算案例的两两相似度,相似度计算公式为:

$$s_{ij}=\frac{\sum_{k=1}^{m} w_{ik}\times w_{jk}D(k)}{\sqrt{\sum_{k=1}^{m} w_{ik}^{2}\sum_{k=1}^{m} w_{jk}^{2}}} \tag{13-24}$$

式(13-24)中的 S_{ij} 为第 i 个和第 j 个案例的相似度值,W_{ik} 为第 i 个案例中 k 个症状的权重,同理 W_{jk} 为第 j 个案例中 k 个症状的权重。$D(k)$为第 i 个和第 j 个案例中第 k 个症状的相似程度,如果两个案例对应的症状相同则 $D(k)$值为 1,不同则为 0。其中症状权重的计算公式为:

$$W_j=P(s_j)\Big/\sum_{1}^{n}P(s_x) \tag{13-25}$$

式(13-25)中的 W_j 为第 j 个症状在这个子案例库中的权重大小；$P(S_j)$为第 j 个症状在此案例库中的总和，而 $\sum P(s)$ 为所有症状的总和。找出最相似且不在同一类的两个案例，如果相似度超过合并的阈值则将两个案例合并为一类，并进入下一次循环，直到不再超过阈值，则分类完成并且程序退出。具体的案例分类算法流程图如图 13-4 所示。分到同一类中的案例一定是类似的，但不一定是同一种疾病案例。分类完成之后，按照公式(13-25)分别对每个子案例库计算症状权重大小，分类之后的案例库结构如图 13-5 所示。

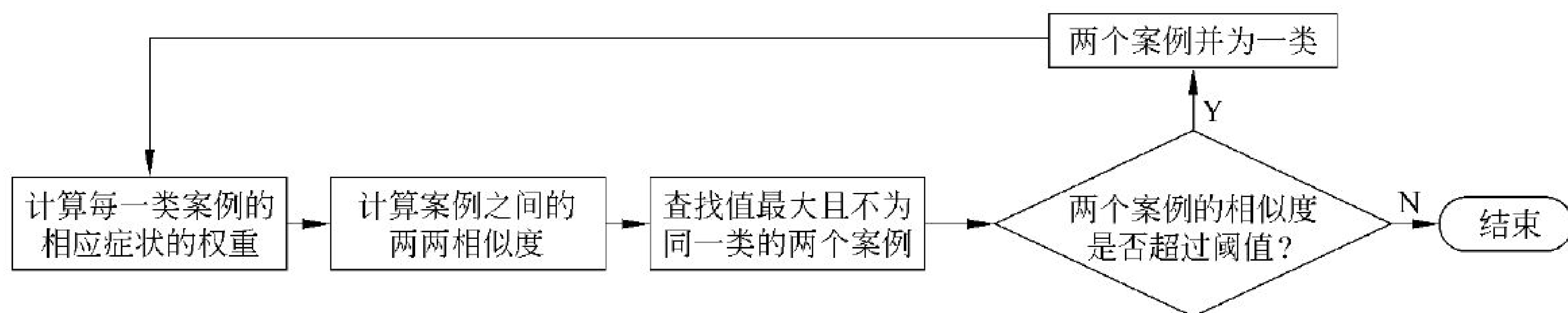


图 13-4　牙鲆疾病案例分类流程图

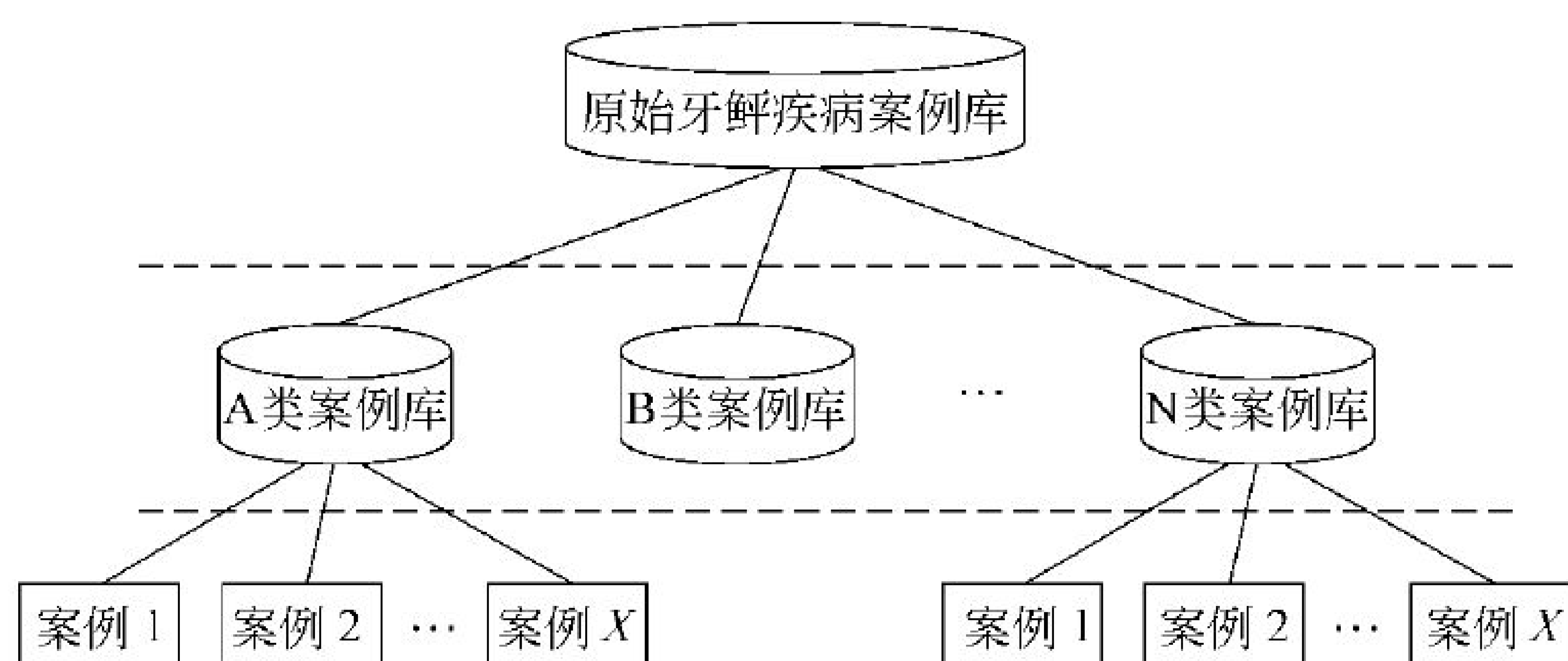


图 13-5　牙鲆疾病案例库结构图

(3) 案例检索与复用

基于案例的牙鲆疾病诊断算法是用旧的案例或经验来分析和解决新的问题。诊断过程是根据用户所输入的症状信息，从牙鲆疾病案例库中寻找与之最相似的案例，并计算得出相似度数值最大的案例。如果相似度值大于诊断阈值(诊断阈值是诊断成功的最低相似度值，根据与专家的讨论研究将其定为 0.65)，则说明诊断成功。若未超过阈值则说明诊断失败，可以让用户进行症状的重新选择或者退出系统。

对于用户每次所输的症状组合来说，都可以将其看成一个新的诊断案例。根据不同的情况，对此案例会有不同的处理方式。若没有找出与此案例的相似度超过阈值的案例，即诊断失败，则将此新案例存入诊断失败的案例库，等待专家进行再次诊断后，将其存入案例库。

若能够找到相似度超过阈值的案例，将存在两种处理方式。当相似度很大时(超过 0.95)，即此案例与案例库中的某一案例基本相同，说明它并没有包含新的诊断信息，因此没有必要保存此案例；反之则说明这个案例有用，并将其存入案例库。经过一定次数的案例复用后，案例库的内容将不断改变，这会影响不同子案例库的权重大小，因此还需要对整个案例库进行重新分类的操作。牙鲆疾病的案例诊断流程如图 13-6 所示。

13.3.2　症状预警警级确定

由上节所述的方法，可以根据牙鲆的发病症状诊断疾病。确定牙鲆所患疾病后，就能从知



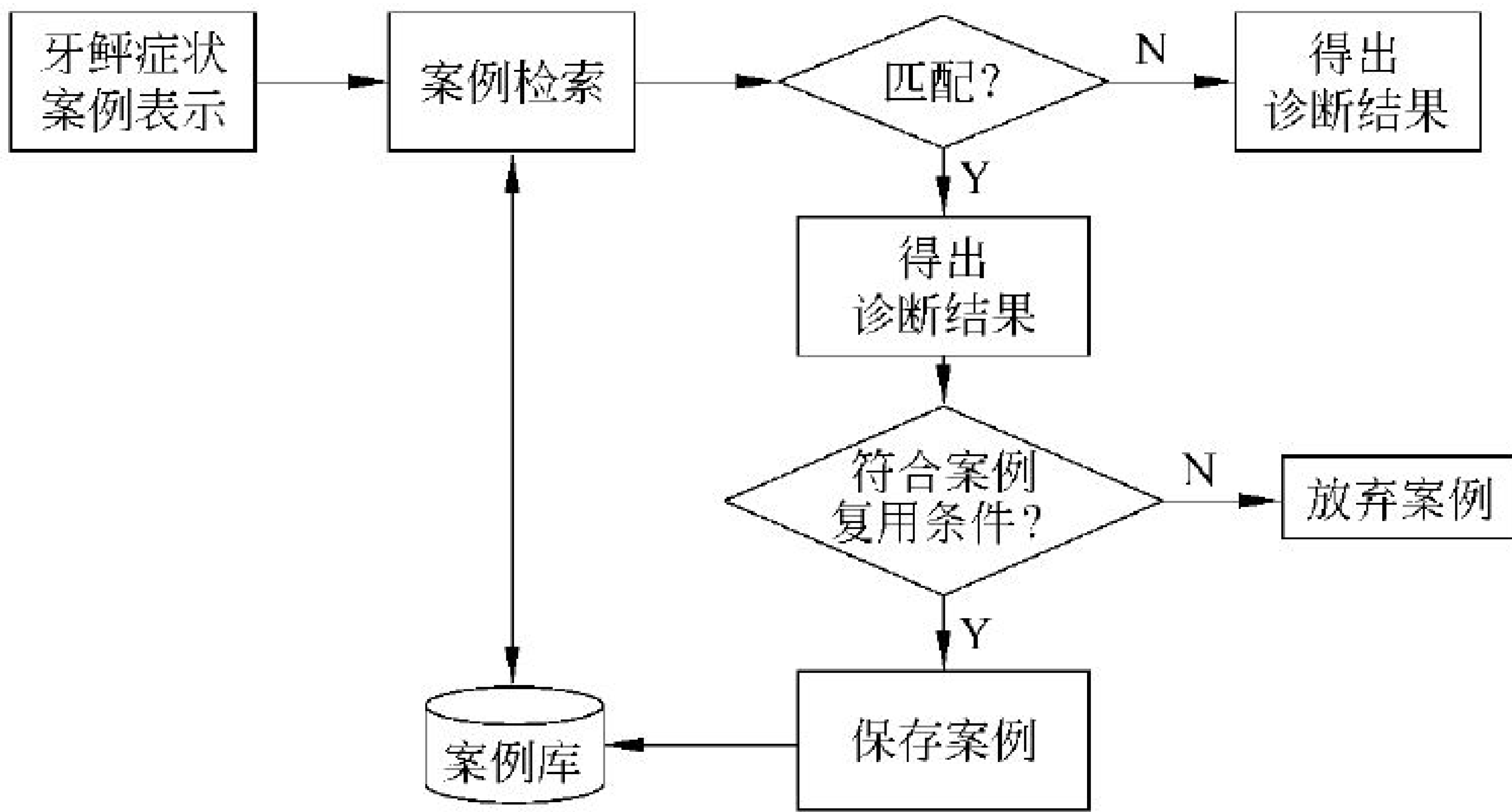


图 13-6　基于案例的牙鲆疾病诊断流程

识库中搜索出该疾病的死亡率和发病率所对应的症状预警的警级大小。其中死亡率和发病率与疾病的对应关系是采用专家调查法获得的，部分数据记录如表 13-10 所示。

表 13-10　症状预警知识表

病　名	无　警		轻　警		中　警		重警和巨警	
	发病率/%	死亡率/%	发病率/%	死亡率/%	发病率/%	死亡率/%	发病率/%	死亡率/%
淋巴囊肿病	0	0	0～1.4	0～0.3	1.4～4	0.3～0.5	>4	>0.5
肠道白浊病	0	0	0～1	0～0.3	1～5	0.3～0.5	>5	>0.5
腹水病	0	0	0～1	0～0.3	1～5	0.3～0.5	>5	>0.5
链球菌病	0	0	0～1	0～0.1	1～5	0.1～0.4	>5	>0.4
出血性败血病	0	0	0～0.8	0～0.1	1～3	0.1～0.2	>3	>0.2
气胞单菌症 I	0	0	0～1	0～0.5	1～6	0.5～2	>6	>2
白点病	0	0	0～1.5	0～0.6	1.5～4	0.6～1.1	>4	>1.1
车轮虫	0	0	0～2	0～0.2	2～5	0.2～0.5	>5	>0.5
盾纤类纤毛虫病	0	0	0～1.5	0～0.4	1.5～5	0.4～0.6	>5	>0.6
淀粉卵涡鞭虫病	0	0	0～1	0～0.1	1～7	0.1～0.4	>7	>0.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

13.3.3　实例分析

我们已针对牙鲆疾病症状预警的基本过程作了详细分析与阐述，下面进行实例分析，包括症状的处理和匹配、案例诊断、案例复用和警级确定等方面。

设用户输入的牙鲆疾病症状为：体色变黑、体表发红有外伤、摄食不良、腹部膨胀、流动迟缓，发病率为 2%，死亡率为 0。首先系统通过从数据库的症状表中搜索到与用户所输入症状对应的编号 ID，并将症状编号从小到大排列，经搜索后得到症状集 casesets 为：1-13-36-39-50。

先将此症状集进行字符串拆分处理，并分别在数据库中的每个子案例库进行相似度分析，最后找出每个子案例库中相似度值最大的案例。再将这些案例的相似度值进行比较，找出相似度值最大一个案例。经计算得出第 74 个案例(病名为盾纤类纤毛虫病，相似度值为 0.783)相似度值最大。

由于此相似度大于0.65,因此诊断成功并将此案例的病名、治疗方法通知用户。由于相似度小于0.95,说明此案例与所匹配的案例并不冲突,因此将用户输入的案例加入案例库。根据诊断出的疾病和用户输入的发病率和死亡率从症状预警知识表中查得属于中警警级,最后将最终警级的大小和预警预案通知用户。系统内部加入用户输入的新案例,并再次进行案例的分类和症状权重的计算。

13.4 小结

本章通过对牙鲆疾病预警指标的分析,确定了以传染源和环境因素为主的警源指标体系,以疾病症状为核心的警兆指标体系,以宏观发病率和死亡率表现为核心的警情指标体系。根据专家调查法对预警指标的重要程度进行排序,选取了溶氧、水温、发病症状、氨氮、pH值等8个预警指标作为最终研究对象,并确定了各个预警指标的无警、中警、重警警限。利用专家知识、神经网络与遗传算法结合的方法,建立了牙鲆疾病的水环境预警模型。利用基于案例的诊断方法对牙鲆的常见疾病进行诊断,并根据疾病名称与发病率的关系确定了牙鲆疾病症状预警警级大小。

参考文献

1. 王瑞梅. 池塘水质管理智能决策支持系统[D]. 中国农业大学,2003
2. Jiawei Han, Micheline Kamber. Data mining concepts and techniques [M]. 2007: 234～238
3. Soman KP, Shyam Diwakar, Ajay V. Insight into data mining theory and practice[M]. 2009: 51～56
4. 国嘉,王瑞敏,王家海,张忠民. 基于遗传算法的神经网络学习方法[J]. 计算机与数字工程, 2004. 32: 98～101
5. 朱勇,吴涛. 基于Rough集和构造性学习神经网络的经济预警模型[J]. 合肥工业大学学报, 2007. 30(7): 836～840
6. 李丽. 基于径向基网络和支持向量机的梨病虫害预警预报系统研究[D]. 中国农业大学,2007.
7. TaeYoon Kima, Kyong Joo Oh, Insuk Sohn, Changha. Hwang. Usefulness of artificial neural networks for early warning system of economic crisis [J]. Expert Systems with Applications, 2004. 26: 583～590
8. 温继文. 基于知识的鱼病诊断推理系统研究[D]. 中国农业大学,2003
9. 郭永洪. 鱼病知识与诊断推理集成[D]. 中国农业大学,2005

应用篇

第 14 章

基于 J2EE 的鱼病诊断系统

本章以鱼病诊断问题为应用背景，采用 J2EE 的体系结构，运行环境为 Windows2000 Server 操作系统，应用服务器为 BEA 公司的 Weblogic 平台，数据库采用 SQL SERVER2000，实现了基于本体的鱼病诊断系统的基本功能，为鱼病诊断系统提供了一个集成的智能化支撑环境。

14.1 鱼病诊断系统框架

14.1.1 J2EE 概述

J2EE (Java 2 Platform Enterprise Edition)是由 Sun 公司于 1999 年 6 月推出的一种利用 Java 2 平台来简化诸多与多级企业解决方案的开发、部署和管理相关的复杂问题的体系结构，是使用 Java 构建企业系统的标准平台和环境。J2EE 企业网络应用程序架构改进了传统 Client/Server 结构的缺点，降低客户端程序逻辑的复杂性，减轻了后台数据库存取的负载，平衡了整个网络的流量，可以使用不同生产商的底层企业系统框架实现。[1]J2EE 不仅巩固了标准版中的许多优点，例如“编写一次、随处运行”的特性、方便存取数据库的 JDBC API、CORBA 技术以及能够在 Internet 应用中保护数据的安全模式等，还提供了对 EJB、Java Servlets API、JSP 以及 XML 技术的全面支持。

14.1.2 J2EE 的优势

J2EE 体系结构提供中间层集成框架来满足无需太多费用而又需要高可用性、高可靠性及可扩展性的应用需求。通过提供统一的开发平台，J2EE 降低了开发多层应用的费用和复杂性，同时提供了对现有应用程序集成强有力的支持，完全支持 Enterprise JavaBeans，有良好的向导支持打包和部署应用，添加目录支持，增强了安全机制，提高了性能。[2]

J2EE 的优势主要表现在以下方面：

(1) 保留现存的 IT 资产。由于市场竞争日趋激烈，企业必须适应用户的各种新需求，利用已有的企业信息系统方面的投资而不是重新制定全盘方案就变得很重要，一个以渐进方式建立在已有系统之上的服务器端平台机制是企业所需求的。J2EE 架构可以充分利用用户原有的投资，而且其拥有广泛的业界支持。由于基于 J2EE 平台的产品几乎能在任何操作系统和硬件配置上运行，现有的操作系统和硬件也能继续使用。

(2) 高效的开发。J2EE 允许公司把一些通用的、很繁琐的服务器端任务交给中间件供应商去完成，这样开发人员可以集中精力于如何创建业务逻辑，相应地缩短了开发时间。

(3) 支持异构环境。J2EE 能够开发部署在异构环境中的可移植程序。基于 J2EE 的应用程序不依赖任何特定的操作系统、中间件、硬件,因此设计合理的基于 J2EE 的程序只需要开发一次就可部署到各种平台。J2EE 标准允许客户订购与 J2EE 兼容的第三方现成组件,把它们部署到异构环境中,节省了由自己制订整个方案所需要的费用。

(4) 可伸缩性。企业必须选择一种服务器端平台,这种平台能够提供极佳的可伸缩性去满足系统上运作的各种新客户。基于 J2EE 平台的应用程序可被部署到各种操作系统上,而且提供了更为广泛的负载平衡策略,能消除系统中的瓶颈,允许多台服务器集成部署。这种部署可达数千个处理器,从而实现可高度伸缩的系统,以满足应用的需要。

(5) 稳定的可用性。一个服务器端平台必须能长时间运转以满足公司客户、合作伙伴的需要。J2EE 部署到可靠的操作环境中,支持长期的可用性。

14.1.3 J2EE 的应用程序结构

J2EE 典型的结构分为以下 4 层。

(1) 运行在客户端机器上的客户层组件。这一层对应于 J2EE 应用程序的客户端。J2EE 客户端可以是基于 Web 方式的,也可以是基于传统方式的,还可以是移动设备。

(2) 运行在 J2EE 服务器上的 Web 层组件。J2EE Web 层组件可以由 JSP 页面或 Servlets 组成,按照 J2EE 规范,静态的 HTML 页面和 Applets 不算是 Web 层组件。Web 层可能包含某些 JavaBean 对象处理用户输入,并把输入发送给运行在业务层上的 Enterprise Bean 进行处理。

(3) 运行在 J2EE 服务器上的业务逻辑层组件。在 J2EE 架构中,业务层组件主要是指 EJB。业务层代码的逻辑用来满足业务领域的需要,由运行在业务层上的 EJB 进行处理。EJB2.0 规范中定义了会话 Bean、实体 Bean 和消息驱动 Bean 三种 EJB。

(4) 运行在 EIS 服务器上的企业信息系统层软件。企业信息系统层包括企业基础建设系统。

J2EE 提供了一个框架用来开发分布式的企业级应用。J2EE 架构中的各项技术需要一个实现者,这个实现者要按照 J2EE 的规范提供应用程序所需要的服务,主要是 Web 服务和业务组件服务,J2EE 应用服务器就是充当这个 J2EE 技术标准实现者的角色。BEA 公司的 Weblogic Server 作为 J2EE 应用服务器,实现了 J2EE 的全部特性。它可以处理一些常见的编程任务,包括提供事务处理服务、安全认证服务、命名和目录服务、数据库访问和连接池、线程池、负载均衡和容错等。[2]

14.1.4 鱼病诊断推理系统的结构

本系统采用 J2EE 体系结构,如图 14-1 所示,操作系统为微软公司的 Windows 2000,应用服务器采用 BEA 公司的 Weblogic,数据库采用 SQL 2000,诊断推理子系统通过 JDBC 数据接口与后台数据库进行数据交互。其中诊断推理子系统部分利用 Borland JBuilder 9 Enterprise 平台进行开发,Web 层组件采用了 JSP 与 Severlets 技术。诊断推理子系统在进行鱼病诊断时根据用户输入的特征属性,与存储的案例进行相似性比较,提取与源案例最相似的案例进行诊断。学习模块则是根据已积累起来的鱼病诊断规则知识,进行规则学习,使用系统自动获取规则。而集成模型则提供了将异构系统进行转换的接口。数据库包括规则库、案例库、案例特征权重库、案例临时库以及案例特征库等。



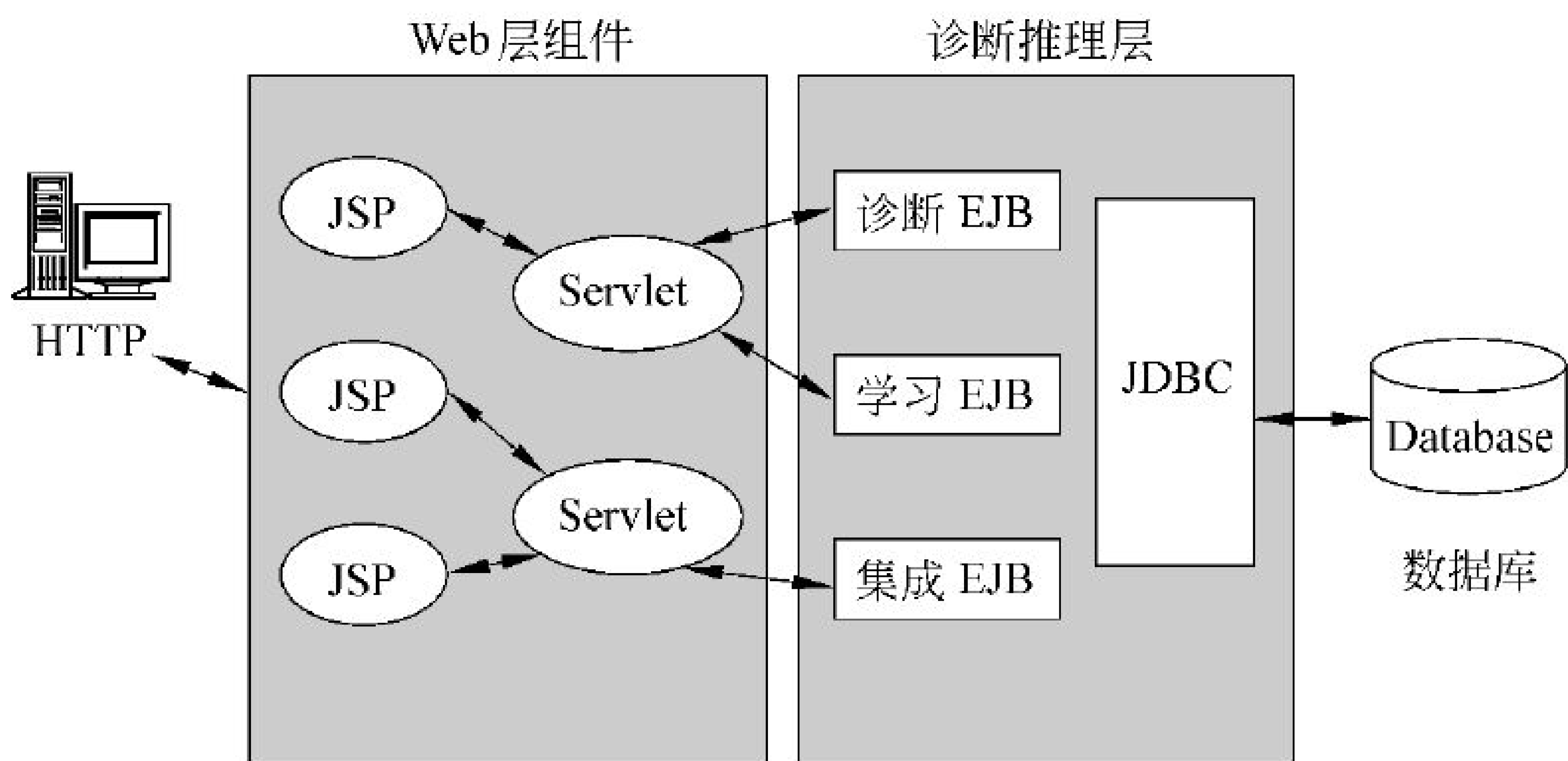


图 14-1　系统体系结构图

本系统在结构上分为三部分：第一部分为基于本体的概念模块，第二部分为规则获取模块，第三部分为基于案例的鱼病诊断推理模块。参见图 14-2。

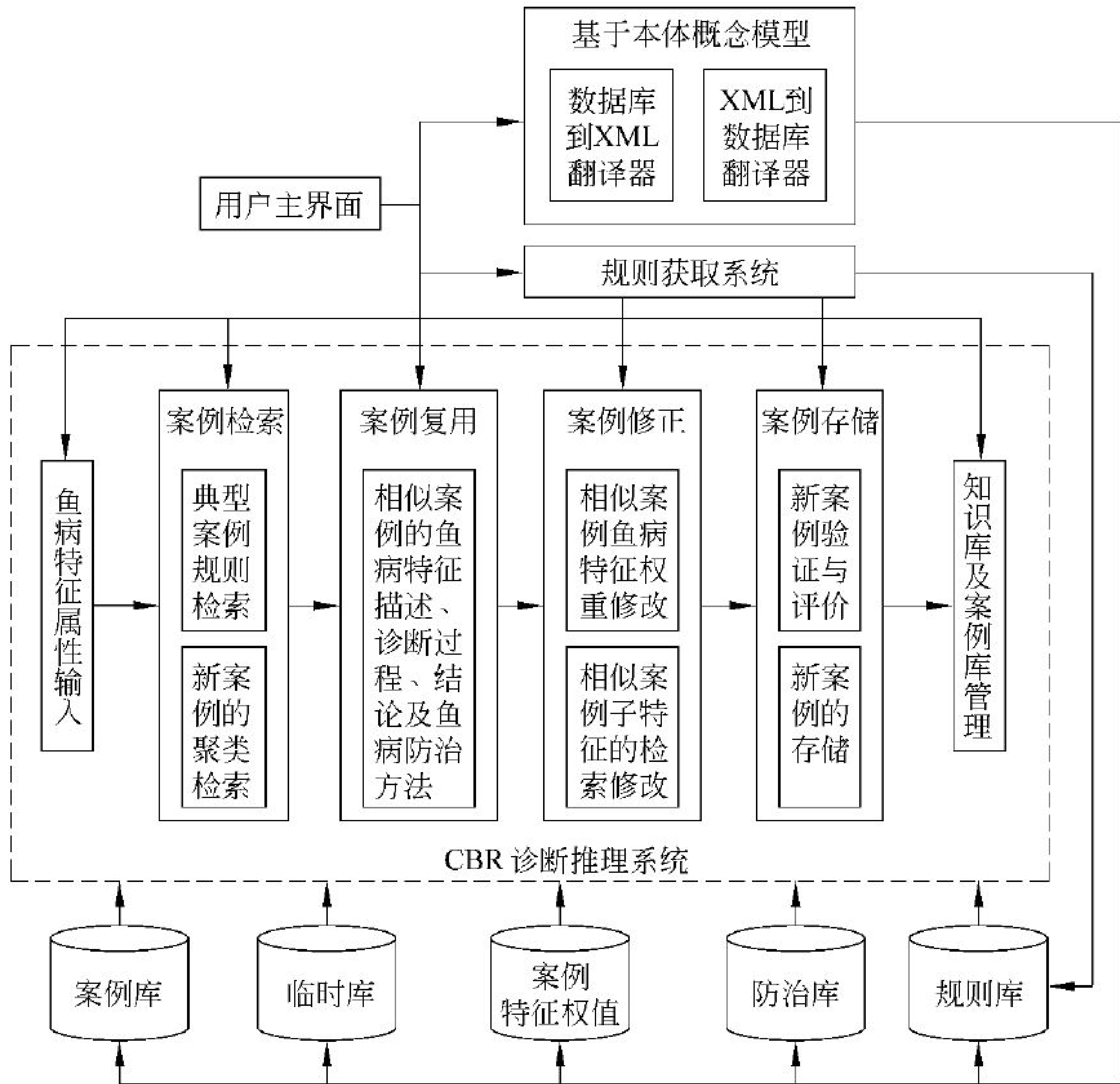


图 14-2　系统结构图

14.2　鱼病诊断流程设计

鱼病种类繁多，发生的病因也很复杂，鱼病发生后是否能尽快地得到控制，对鱼病迅速作出正确诊断是首要步骤。只有首先确诊，找出致病原因，才能对症下药，取得控制疾病的效果。因此，能否对鱼病及时作出正确诊断，是防治鱼病中的关键问题[3]。鱼病诊断的步骤和方法如下所述。

(1) 调查询问。在诊断某种鱼病之前,首先向渔农调查询问,这一点很重要,因为它可以得到现场无法得到的材料。涉及的内容主要包括是单养还是混养,混养鱼类的死亡顺序以及放养的密度、施肥情况、当时的天气变化和死鱼前浮头的情况,此外,还包括以前治疗情况,如用过何种药物、使用药物浓度和次数、效果怎么样。

(2) 现场调查。鱼生病后,单纯对鱼体进行观察,有时很难作出正确诊断,现场观察也是诊断鱼病的重要环节。现场观察包括水质状态、鱼在水中的表现、鱼池的周围环境等内容。首先要了解水质变化。水质的透明度、肥瘦情况以及水质的酸碱度和溶氧情况都是可使鱼类生病的因素。必要时可进行测定,特别是对一些病症不明显较难以确诊的鱼病。其次,了解病鱼在水中的表现。通常沿池塘巡视一周,观察濒死鱼的游姿和体色的改变等。急性型鱼病,病鱼的体色和正常鱼差不多,仅在病变的部位稍有变化,但一经死亡,死亡率随即上升,常在短期内出现死亡高峰。而慢性型鱼病则往往体色变黑,体质瘦弱,离群独游,死亡率一般上升缓慢,在相当长的时间内出现死亡高峰。有些病鱼一旦离开水体或死后,一些极其重要的症状就看不到了,给确诊造成困难。如草鱼白头白嘴病,当病鱼在水中时,症状很明显;离开水体后,白头白嘴现象就看不到了。而根据这些特别症状很容易作出诊断。

(3) 鱼体检查:经调查询问和现场观察,对鱼病发生的有关外部环境因素有了大体的了解,对某些鱼病可能会作出诊断。但对大多数鱼病要作出正确的诊断,则主要靠鱼体检查。检查病鱼时,最好捞取濒临死亡而未死亡的病鱼进行检查,如果达不到这一要求,也要尽可能选取刚死亡但尚未腐败的病鱼进行检查。鱼体检查首先是目检,目检是检查鱼病的主要方法之一,有些鱼病仅通过肉眼检查就可确诊,由于病原体的寄生部位往往呈现出一定的病理变化,有时症状很明显。如水霉以及一些大型的寄生虫,用肉眼可能识别出来。但有些病原体(如细菌、体型较小的寄生虫),用肉眼是看不见的。但是一般细菌性鱼病,常表现出各自的不同症状(充血、发炎、脓肿、腐烂、蛀鳍等症状);而寄生虫性鱼病,常表现出黏液增多、发白、有点状或块状的胞囊等症状。通过肉眼观察其不同的症状,对于某些鱼病就可作出初步诊断。所以肉眼检查法是一种较为方便并有一定效果的方法。对鱼体检查时,主要检查体表、鳃、内脏三大部分。肉眼检查主要是以症状为依据,如果同一尾鱼体并发两种以上的病,同时有几种病的症状,就很难确定患何种病,如体色发黑、蛀鳍、烂尾、鳃丝分泌黏液增多等症状,在这种情况下,仅靠肉眼检查是不能确诊的,必须进一步借助显微镜或解剖镜,方可作出较正确的诊断。

(4) 显微镜检查(镜检)。对一般肉眼不能看到的病鱼症状或病原体,必须借助于显微镜或解剖镜进一步检查。镜检一般先要用目检来确定病变部位,然后再用显微镜作细微的全面检查。

调查询问、现场调查、鱼体检查是诊断鱼病缺一不可的整体步骤,只有将各方面所得到的材料进行综合分析,去粗取精、去伪存真,才有可能作出正确的诊断。

了解了鱼病专家对鱼病诊断的实际情况后,我们采用基于案例推理的诊断方法,得出鱼病诊断的推理过程与 CBR 系统的推理过程是一致的,即新案例的描述、案例检索、案例复用、案例修正与评价及案例存储等几个过程。图 14-3 是基于案例推理的鱼病诊断系统的流程图。

图 14-3 中的具体步骤如下所述。

(1) 鱼病案例的输入。采用基于特征的诊断方法,对新案例进行信息描述,人工交互式地输入询问调查、现场调查及目检观测到的特征属性,如所观测到的各种症状、环境等特征



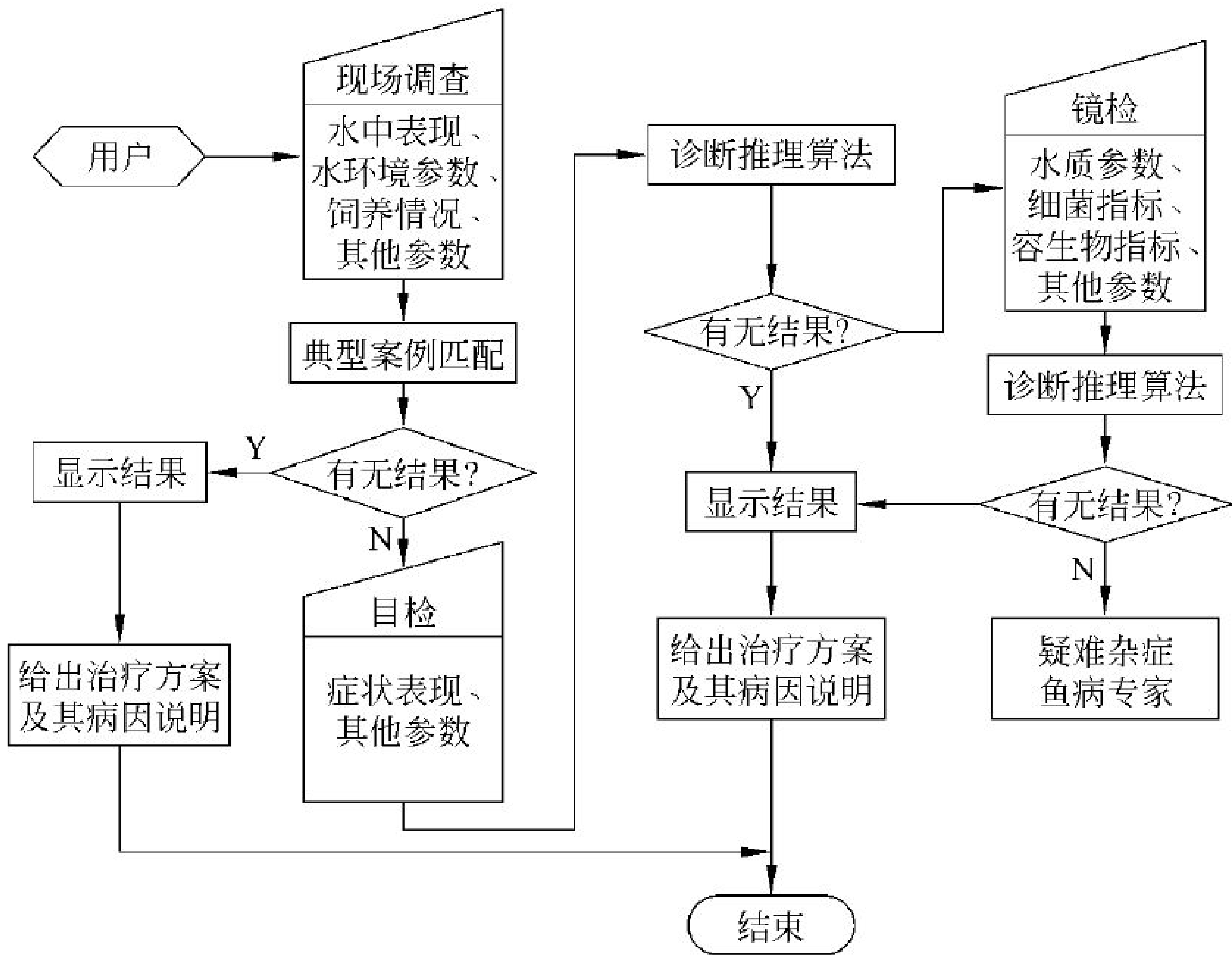


图 14-3　鱼病诊断推理流程

信息。

(2) 相似病例的检索。相似案例的检索是基于案例推理的一个十分重要的环节,检索结果直接影响推理的成功与否。根据新案例中鱼的种类,检索模块从相应的规则库中取出鱼病案例,并逐一与新案例进行相似性比较,根据相似算法,计算出相似度值并排列出相似病例的清单。

(3) 规则案例复用。一旦检索模块找出了相似病例,则从病例库中调出该病例,初步确定鱼所患的疾病。

(4) 疾病诊断修正。检索出的相似病例一般并不能直接作为新案例的诊断结果,由于检索出的相似病例的属性和特征与新案例的属性和特征并不完全相同,检索出的相似病例只是与新案例在总体上相似,所以存在差异是必然的,这些差异包括症状所表现的部位并不完全相同。检索出的相似案例必须经过适当修正,才能适用于新案例,这些修正包括对案例某些症状的添加或删除,它是 CBR 子系统的一个重要环节。[4~6]

(5) 疾病案例的存储。修正后的病例需要经过审核,如果该案例通过审核并在以后的实际诊断中被证明是可行的,具有一定的参考价值,则将该病例存入到案例库中去,以备以后的鱼病诊断使用。

14.3　数据模型设计

14.3.1　规则库和案例库的设计

1. 规则库由条件和结论两部分组成

表示当条件满足时可得出相应结论。规则库中规则表的设计如表 14-1 所示。

表 14-1 规则表

字段名称	字段含义	数据类型	长度
RuleID	规则编号	Varchar	3
Condition 1	满足条件 1	Varchar	4
Condition 2	满足条件 2	Varchar	4
⋮	⋮	⋮	⋮
Condition n	满足条件 n	Varchar	4
Result	结论	Varchar	10
ClassID	规则分类标识	Varchar	3

2. 案例库的设计

现场调查是鱼病诊断的第一步，现场观察的内容包括水质状态、鱼在水中的表现、鱼池的周围环境等内容，对其主要特征进行归纳可总结出事实表，案例库中事实表的设计如表 14-2 所示。

表 14-2 现场诊断事实表

字段名称	字段含义	数据类型	长度
CaseID	案例编号	Varchar	3
ObjectID	诊断对象编号	Varchar	4
Behave	表现描述	Varchar	4
DiseaseID	疾病编号	Varchar	4
DisCause	病因说明	Varchar	10
CureSuggest	治疗建议	Varchar	20

3. 案例权值数据表

案例权值数据表用于储存案例中各特征所对应的权值，表中每一条记录存储了一种疾病的各部位症状的权值，并代表一个案例。这两个表中的记录通过“DiseaseID”字段实现一一对应关系。

表 14-3 中，一条记录对应的是某种疾病反映在鱼体各个部分的症状和鱼所处环境特征的权值，字段 $w_1, w_2, \cdots, w_n$ 用于记录各特征的权值，这些字段与系统案例集合中的最大特征集合中的元素相对应，其个数也等于最大特征集合的个数。

表 14-3 案例权值数据表

字段名称	字段含义	数据类型	长度
DiseaseID	疾病编号	Varchar	3
DiseaseName	疾病名称	Varchar	10
w_1	权值 1	Varchar	4
w_2	权值 2	Varchar	4
⋮	⋮	⋮	⋮
w_n	权值 n	Varchar	4



4. 案例特征表

在该表中，每条记录对应一个案例。"疾病 ID"是主索引字段，并且该字段与案例权值数据表中"疾病代码"字段相对应，二者共同完成对案例的标定。字段 $F_1,F_2,\cdots,F_n$ 代表一个案例特征，用于记录各特征在不同案例中的特征值，这些字段与系统案例集合的最大特征集合中的元素相对应，其个数也等于最大特征集合中元素的个数。

表 14-4　案例特征值数据表

字段名称	字段含义	数据类型	长度
DiseaseID	疾病编号	Varchar	3
DiseaseName	疾病名称	Varchar	10
F_1	特征值 1	Varchar	4
F_2	特征值 2	Varchar	4
⋮	⋮	⋮	⋮
F_n	特征值 n	Varchar	4

14.3.2　本体模块的设计

前面已经阐述了本体设计的方法及本体框架，[7~9]在此给出病因、治疗及方法的本体的详细设计。见图 14-4 至图 14-6。

```
<?xml version=1.0 encoding="GBK" standalone=" yes"?>
<Defontology>
  <ontologyName>病因本体</ontologyName>
  <Body Section>
        <属性 1>
            <名称>最常见病因</名称>
            <类型>字符串数组</类型>
            <例子>最常见致病因素、最常见致病因子、最常见诱因、最常见致病药物、最
            常见致病食品</例子><注释>常见致病的因子、药物、食物等</注释>
        </属性 1>
        <属性 2>
            <名称>常见病因</名称>
            <类型>字符串数组</类型>
            <例子>常见致病因素、常见致病因子、常见诱因、常见致病药物、常见致病食
            品</例子>
            <注释>常见致病的因子、药物、食物等</注释>
        </属性 2>
        <属性 3>
            <名称>少见病因</名称>
            <类型>字符串数组</类型>
            <例子>少见致病因素、少见致病因子、少见诱因、少见致病药物、少见致病食
            品</例子>
            <注释>少见致病的因子、药物、食物等</注释>
```

图 14-4　病因本体

```
        </属性 3>
        <属性 4>
            <名称>罕见病因</名称>
            <类型>字符串数组</类型>
            <同义词>偶见病因</同义词>
            <例子>罕见致病因素、罕见致病因子、罕见诱因、罕见致病药物、罕见致病食
            品</例子>
        </属性 4>
        <属性 5>
            <名称>直接病因</名称>
            <类型>字符串数组</类型>
            <例子>直接致病因素、直接致病因子、直接诱因、直接致病药物、直接致病食
            品</例子>
          <注释>直接导致疾病的原因</注释>
        </属性 5>
        <属性 6>
            <名称>病程</名称>
            <类型>字符串数组</类型>
            <例子>年、月、日、星期</例子>
            <注释>从疾病的发生到疾病的转归好转的时间</注释>
        </属性 6>
        <属性 7>
            <名称>主要病因</名称>
            <类型>字符串数组</类型>
            <注释>导致疾病的主要原因</注释>
        </属性 7>
        <属性 8>
            <名称>次要病因</名称>
            <类型>字符串数组</类型>
            <注释>导致疾病的次要原因</注释>
        </属性 8>
        <属性 9>
            <名称>病因</名称>
            <类型>字符串数组</类型>
            <例子>致病因素、致病因子、诱因、致病药物、致病食品</例子>
            <注释>不属于上述病因,则用此属性</注释>
          </属性 9>
</Body Section>
<Relation Section>
    <继承>基础本体</继承>
</Relation Section>
<Documentation Section>
⋮
</Documentation Section>
</Defontology>
```

图 14-4（续）



```
<?xml version=1.0 encoding="GBK" standalone=" yes" ?>
<Defontology>
  <ontologyName>治疗方法本体</ontologyName>
  <Body Section>
        <属性 1>
            <名称>治疗部位</名称>
            <类型>字符串数组</类型>
            <值域>身体部位</值域>
        </属性 1>
        <属性 2>
            <名称>治疗途径</名称>
            <类型>字符串数组</类型>
        </属性 2>
        <属性 3>
            <名称>副作用</名称>
            <类型>字符串数组</类型>
            <注释>指治疗对疾病对象所产生的不良作用</注释>
        </属性 3>
        <属性 4>
            <名称>治疗过程</名称>
            <类型>字符串数组</类型>
        </属性 4>
        <属性 5>
            <名称>治疗效果</名称>
            <类型>字符串数组</类型>
            <值域>良好、好、差、较差、明显、不明显</值域>
        </属性 5>
        <属性 6>
            <名称>治疗成功率</名称>
            <类型>百分数</类型>
            <注释>范围、地区、鱼群、条件、时间</注释>
        </属性 6>
        <属性 7>
            <名称>疗程</名称>
            <类型>数量</类型>
            <单位>年、月、日、星期</单位>
        </属性 7>
        <属性 8>
            <名称>疗效评定</名称>
            <类型>字符串数组</类型>
        </属性 8>
  </Body Section>
  <Relation Section>
        <继承>基础本体,治疗方法</继承>
```

图 14-5　治疗方法本体

```
  </Relation Section>
  <Documentation Section>
  ⋮
  </Documentation Section>
</Defontology>
```

图 14-5（续）

```
<?xml version=1.0 encoding="GBK" standalone=" yes" ?>
<Defontology>
  <ontologyName>治疗本体</ontologyName>
  <Body Section>
        <属性 1>
            <名称>治疗原则</名称>
            <类型>字符串数组</类型>
            <注释>指治疗疾病时应遵从的原则</注释>
        </属性 1>
        <属性 2>
            <名称>治疗依据</名称>
            <类型>字符串数组</类型>
            <同义词>治疗根据</同义词>
            <注释>指治疗疾病时的依据</注释>
        </属性 2>
        <属性 3>
            <名称>治疗方法</名称>
            <类型>字符串数组</类型>
            <注释>指治疗疾病时的方法</注释>
        </属性 3>
        <属性 4>
            <名称>早期治疗方法</名称>
            <类型>字符串数组</类型>
            <注释>指病症早期应采用的治疗方法</注释>
        </属性 4>
        <属性 5>
            <名称>晚期治疗方法</名称>
            <类型>字符串数组</类型>
            <注释>指病症晚期应采用的治疗方法</注释>
        </属性 5>
        <属性 6>
            <名称>痊愈</名称>
            <类型>字符串</类型>
            <注释>指疾病痊愈后好转的预期估计情况</注释>
        </属性 6>
        <属性 7>
```

图 14-6　治疗本体



```
            <名称>转归</名称>
            <类型>字符串</类型>
            <注释>指疾病最终的转化发展</注释>
        </属性 7>
        <属性 8>
            <名称>预防方法</名称>
            <类型>字符串数组</类型>
            <同义词>预防措施、预防手段</同义词>
            <注释>预防疾病发生发展的方法和措施</注释>
        </属性 8>
        <属性 9>
            <名称>禁忌</名称>
            <类型>字符串</类型>
            <注释>指在治疗疾病时的禁忌事项</注释>
        </属性 9>
    </Body Section>
    <Relation Section>
        <继承>基础本体,治疗方法</继承>
    </Relation Section>
    <Documentation Section>
    ⋮
    </Documentation Section>
</Defontology>
```

图 14-6（续）

14.4　系统功能及界面设计

14.4.1　功能设计

基于案例推理的鱼病诊断系统具有基于本体的概念模型、案例输入、系统输出、规则学习、权值学习、鱼病诊断推理等功能模块。

基于本体的概念模型：该模型中，鱼病诊断推理子系统的知识表示语言及知识库是用数据库实现的；同样，鱼生存环境的水质评价系统中的知识表达及知识库也是通过数据库实现的。而两个数据库的结构是完全不同的，两个系统之间不能直接共享信息，为此，一方面我们将领域本体知识用 XML 表达出来，当诊断推理系统需要时，通过翻译器将其转化到数据库中，另一方面，若推理系统需要某些评价系统的参数时，也可以将数据库中的内容转化为 XML 格式，这样就可以共享领域的事实知识，实现系统的集成。[10] 本模块实现了翻译器的功能，它有两部分的内容，一部分是将 XML 文档转换成数据库文件，另一部分是将数据库文件转换为 XML 文档。

案例输入模块：主要是根据案例的表示中的一些特征属性，向案例库中输入这些特征属性值，首先确定鱼的品种、鱼的规格、发病时间以及身体各部分的症状表现等。

系统输出模块：主要是对于系统中各功能模块所获得的结果进行输出。

规则学习模块：其实现的主要思想是根据示例学习的目标寻找最优覆盖规则，即生成具有最少数目公式覆盖。本设计中利用遗传算法来寻求最优覆盖规则。根据 AQ 算法的思想，结合鱼病诊断知识，我们按如下的方法构造规划模型：第一类约束条件要求第一个正例必须被至少一个原子公式及所代表的逻辑规则所覆盖，第二类约束条件要求除第一个正例之外的全部正例必然被公式所覆盖，第三类约束条件要求全部反例不能被任何公式所覆盖，或者只能被该公式的某些原子公式所覆盖。该模型将概念学习问题转化为一个规划模型。

权值优化模块：案例学习是 CBR 中的关键技术问题之一，它的性能和效率对保证系统对问题求解的有效性、准确性有着至关重要的影响，而案例的特征项权重对检索的质量与速度都起着重要的作用。由于相似度量时案例的各特征属性所起的作用是不同的，而且这种作用的重要性会随着分类的改变而变化。通过对特征权值的调整和优化是案例修正的一个重要方式，同时也是进行案例学习的一种常用的方法手段。而实际应用中，案例特征权值向量难以确定，基于遗传算法的随机全局搜索能力，本模块实现了一种利用遗传算法来完成特征项权空间的直接搜索的算法，由遗传算法进行特征权值的优化。[11]

鱼病诊断推理模块：系统运行时，根据分类结果建立案例库，采用两级索引。当向系统输入新的鱼病诊断症状时，首先在规则集合中进行归类，由此得出和目标案例相似的一类源案例，初步确定疾病。根据初步确定的疾病，在案例库中根据分类标识进行匹配，若完全匹配，确诊疾病，并给出相应的治疗方案，若不完全匹配，则根据分类标识用 KNN 法计算找出和目标疾病相似度最大的案例，并进行适当修正，[12]重新进行检索，该检索在大幅度缩小的搜索空间里进行，检索效率明显提高，这一步实际上是进行的反向推理，在案例库中来证实上一步得出的结论是否正确。

14.4.2 用户输入界面

用户输入界面功能是用户根据调查询问、现场调查、目检等得到的症状、环境参数，进行新案例的录入。鱼病诊断推理系统的输入界面如图 14-7 和图 14-8 所示。

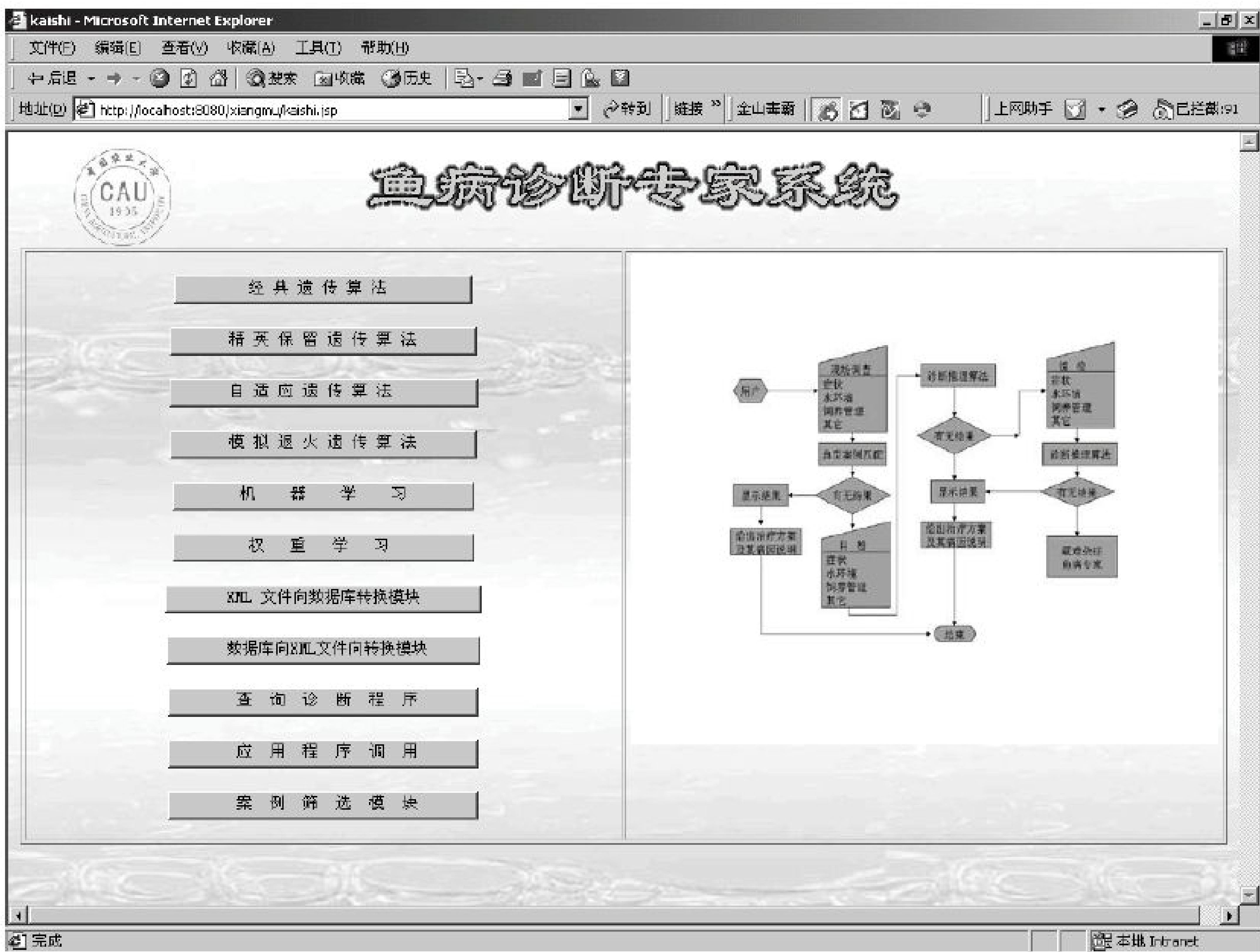


图 14-7 鱼病诊断推理系统的输入界面





图 14-8　鱼病诊断模块输入界面

14.4.3　诊断推理

诊断推理功能包括三部分。第一部分是精确案例匹配，第二部分是基于规则库案例相似检索，第三部分是基于案例库的最大相似检索。鱼病诊断结果界面如图 14-9 所示。

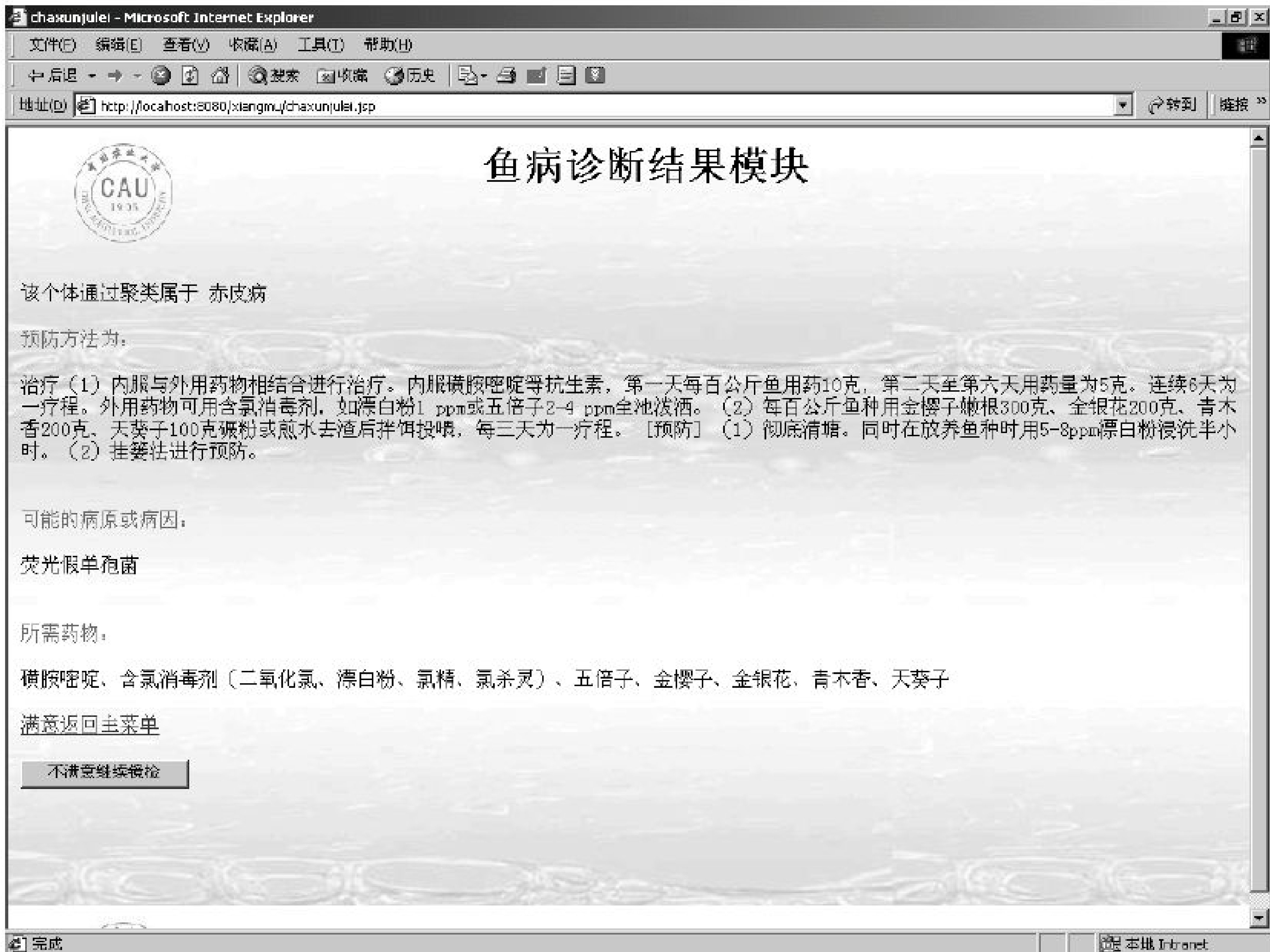


图 14-9　鱼病诊断结果界面

14.4.4　知识获取

知识获取功能是主要是针对规则库，以金鱼烂鳃病病例作为正例，肠炎病病例作为反例，实现了基于 GA 的概念学习模型，学习结果如图 14-10 所示。



图 14-10　机器学习结果显示界面

14.5　小结

本章建立了基于本体设计的鱼病诊断系统原型,并根据问题实例分析了系统运行的情况。主要工作及创新如下:

(1) 建立了基于本体设计的鱼病诊断系统的体系结构。

(2) 系统的设计符合鱼病专家的诊断思想,成功地用 Java 语言实现了推理过程。

(3) 该系统具有自学习功能,用遗传算法解决了权重的分配问题,随着鱼病诊断案例的增加,系统解决问题的能力将逐渐增强。

参考文献

1. 赵强,乔新亮. J2EE 应用开发[M]. 北京:电子工业出版社,2003
2. 李华强,要水根,郭英奎等. 精通 JAVA 中间件编程[M]. 北京:中国水利水电出版社,2003
3. 王振龙,宋憬愚. 鱼病诊断与防治[M]. 北京:中国农业大学出版社,1998
4. 鄢琦. 基于非同构案例集合的 CBR 技术研究及在故障诊断中的应用[D]. 北京科技大学,2002
5. 邓志鸿,唐世谓. 面向语义集成——本体在 Web 信息集成中的研究进展. 计算机应用,2002,22(1):15~17
6. 魏元凤,骆洪青,辛崇波等. 属性相似案例的检索模型比较研究. 华东船舶工业学院学报,1999,13(4):41~44
7. 张凯,王瑜,袁时金. 基于本体集成的资源共享平台. 计算机工程,2003,29(21):59~60
8. 傅谦,张申生,胡锦敏. 基于本体的领域知识共享研究. 计算机集成制造系统——CIMS,2001,7(11):54~58
9. Narasimha Bolloju,Mohamed Khalifa. Integration knowledge management into enterprise environments for the generation decision support[J]. Decision Support Systems,2002(33):163~176
10. 王昕,熊光楞. 基于本体的设计原理信息提取[J]. 计算机辅助设计与图形学学报,2002,14(5):429~432
11. 倪志伟,蔡庆生,贾瑞玉. 范例库中特征项权重的发现技术. 厦门大学学报,2002,41(2):168~172
12. Hori K. An ontology of strategic knowledge: key concepts and applications[J]. Knowledge-Based Systems. 2000,13(6):369~374

第15章

基于案例推理的鱼病诊断专家系统

针对现行鱼病诊断专家系统知识获取与诊断推理的不足,[1]前面理论部分提出从案例推理的角度研究鱼病诊断问题,并综合鱼病诊断系统中案例元素间的关系,给出了基于案例推理的鱼病诊断模型,确定了合适的案例表示方式、存储结构以及鱼病诊断问题求解策略。本章根据专家系统的技术开发要求,以 Jbuilder9+DreamweaverMX+SQL2000 为开发工具,在 Windows2003 Server+Apache+Tomcat5.0 系统/客户开发平台下,采用 JSP+JavaBeans 的应用层模式,建立了基于案例推理的鱼病诊断专家系统。

15.1 系统体系结构设计

基于案例推理的鱼病诊断专家系统是基于 Internet 运行环境开发的,采用当前流行的"Web 浏览器/Web 服务器/数据库系统"网络体系结构(也称 B/S/S 结构),基本思想是在分布式技术基础上,将客户层同应用层分离,组件封装在应用层,系统按功能划分为客户层、应用层和数据层三大块,分别放置在相同或不同的硬件平台上,[2]系统的三层分布式网络体系结构如图 15-1 所示。

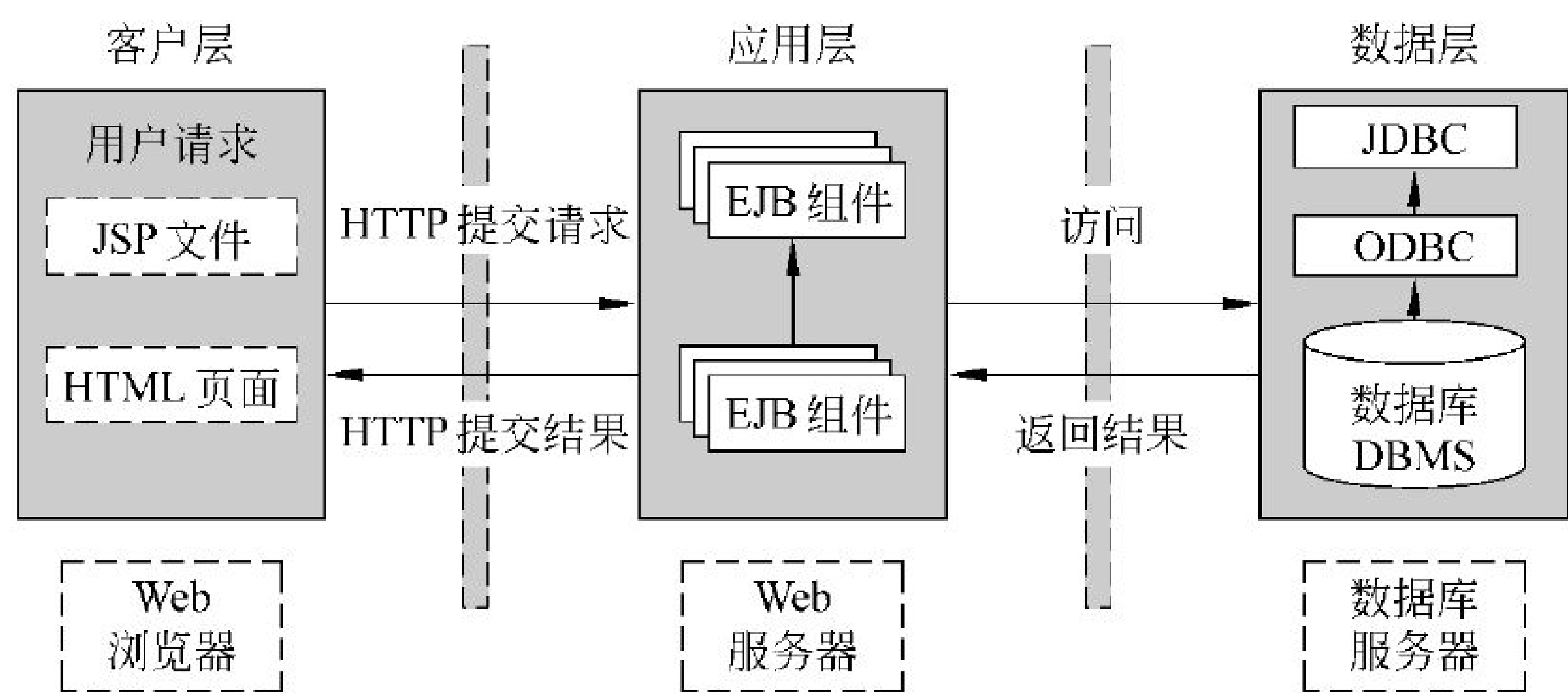


图 15-1　系统体系结构图

(1) 客户层:Web 浏览器为客户层,是系统的用户接口部分,是用户与系统间交互信息的窗口。它的主要功能是指导操作人员使用已定义好的服务或函数,检查用户输入的数据,显示系统输出的数据。

(2) 应用层:Web 服务器为应用层,它在客户层和后端数据库之间操作,也可称为中间业务逻辑层。本系统应用层的主要功能是提供系统中与应用逻辑有关的各种服务构件,负责处理前端客户层的应用请求,它是将原先置于前端客户层的事务逻辑分离出来置于服务器部分,完成事务逻辑的计算任务,并将处理结果返回给用户。在系统开发过程中,利用EJB组件将

应用逻辑封装起来，实现基于案例推理的中间业务逻辑。

(3) 数据服务层：就是数据库管理系统(databases management system，DBMS)，负责提供和管理各类数据，包括数据库、知识库、多媒体库及模型库中各类数据项的访问。该层主要是通过中间业务逻辑层应用逻辑组件为若干个客户共享数据库的连接，向前端客户层、中间应用层提供数据处理与来源，从而减少连接次数，提高数据服务器的性能和安全性。

15.2 系统诊断流程设计

基于案例推理的鱼病诊断专家系统诊断流程分为如图15-2所示的几个步骤。

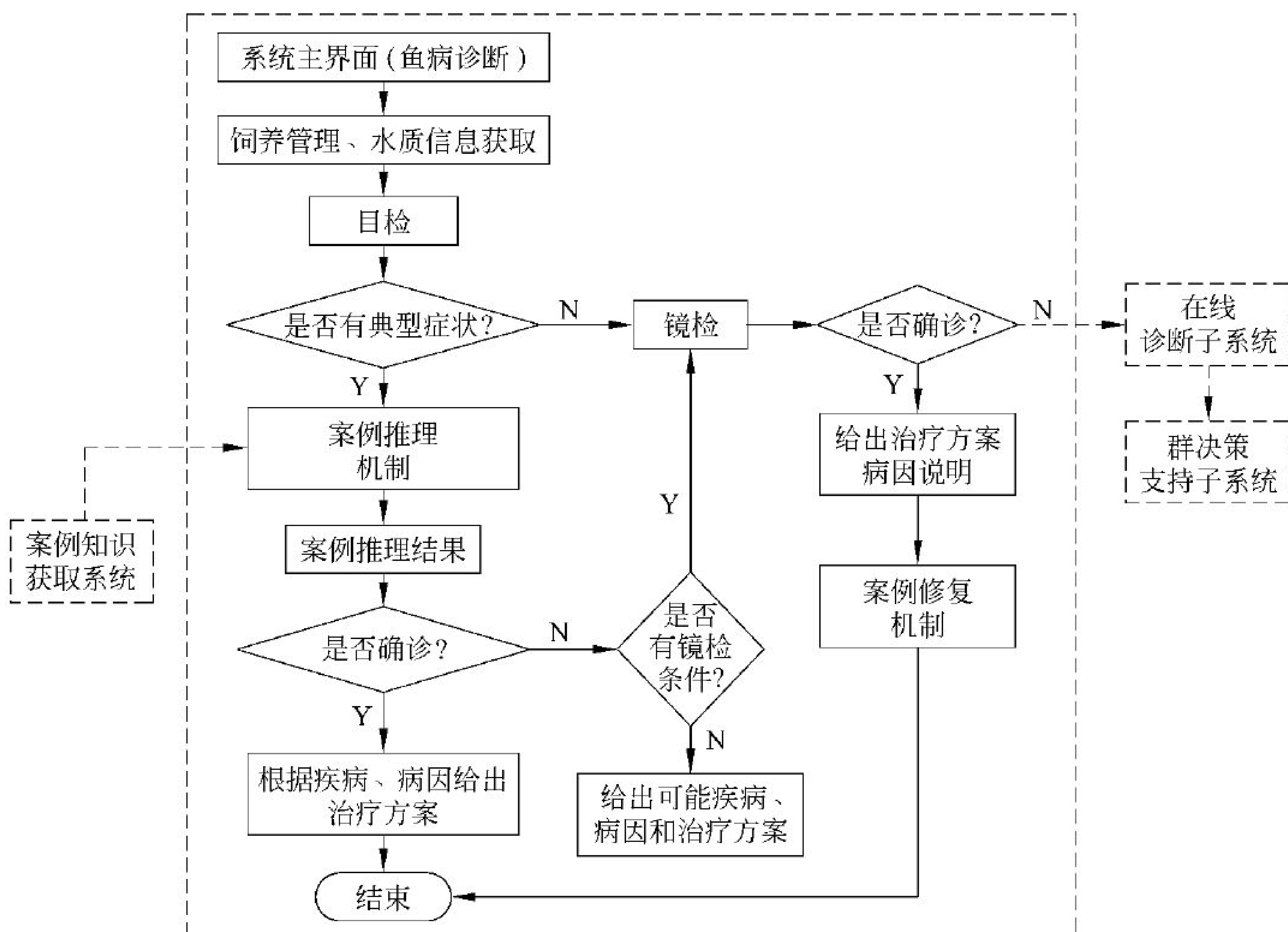


图15-2　鱼病诊断专家系统诊断流程图

(1) 首先案例知识获取系统为鱼病诊断系统获取确诊案例，作为典型案例存入案例库；

(2) 用户通过系统主界面输入饲养管理信息、水质信息、病鱼的池中表现以及目检鱼体以及水体的典型症状，若无典型症状，则直接进入镜检阶段(5)；

(3) 系统根据(2)提供的症状信息进行案例推理；

(4) 如果确诊则根据疾病、病因给出方案，结束诊断；否则，根据案例推理的提示信息进入镜检阶段(5)；

(5) 镜检鱼体和水体，进行深层判断；

(6) 通过镜检，如果确诊，则对案例进行修复，案例修复完毕，结束诊断；否则求助在线诊断系统和群决策系统。



15.3　系统功能模块设计

根据整个系统流程设计，基于案例推理的鱼病诊断专家系统由七大功能模块组成：鱼病案例远程获取模块、水质及鱼体症状信息输入模块、案例诊断推理模块、深层判断模块、鱼病案例修正模块、诊断结果或建议输出模块、系统后台维护模块（如图 15-3 所示）。

鱼病案例远程获取模块：实现对注册专家远程递交的鱼病电子病历的收集与整理，并将数据输入典型案例库。

水质及鱼体症状信息输入模块：用户通过该模块实现新的鱼病案例的水环境信息、饲养管理信息、鱼体症状信息的输入。

案例诊断推理模块：该模块采用案例推理方法模拟专家形象思维，通过对典型案例库的筛选和检索，实现新的鱼病案例的诊断。诊断推理模块包括基于相似性的案例检索模块和基于粗分析的案例检索模块，两个子模块又分别包括三个模块，如图 15-4 所示。

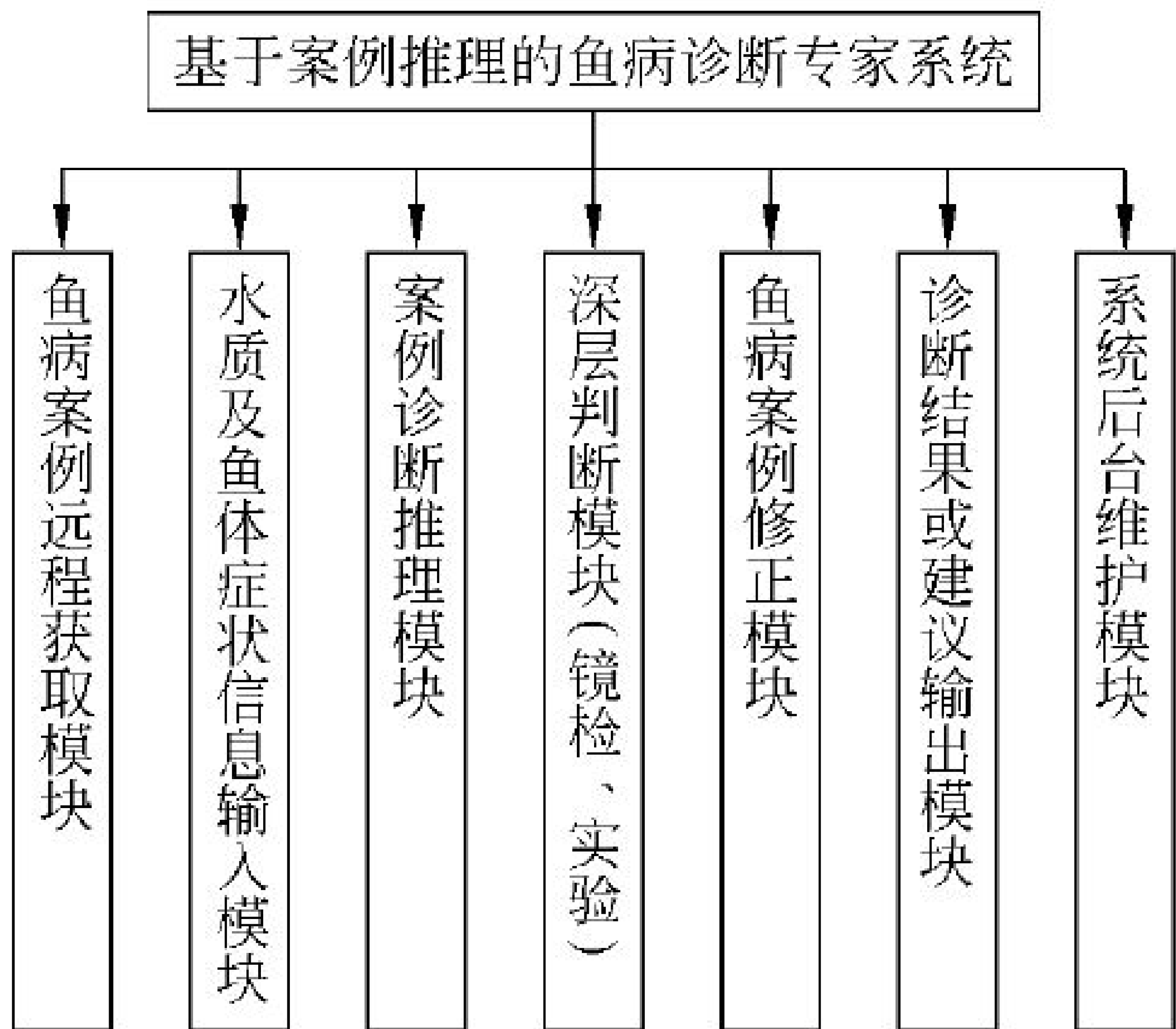


图 15-3　系统功能模块

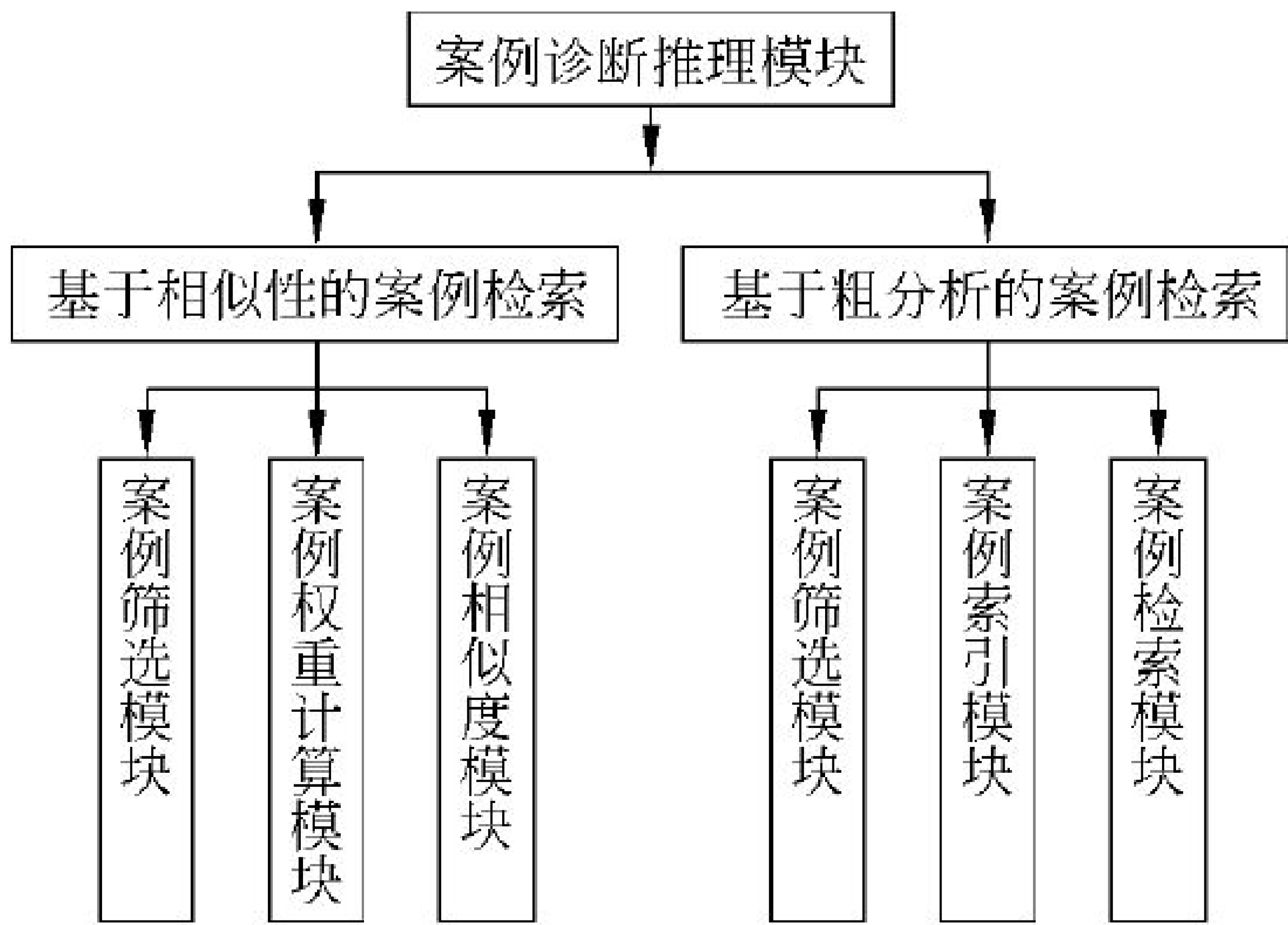


图 15-4　诊断推理模块

深层判断模块：当案例诊断推理模块不能对新的鱼病案例确诊时，进入疑难案例深层判断模块。该模块主要是对病鱼体或水体做进一步的显微镜检查或微生物鉴定试验，为不常见的鱼病做深度判断。

鱼病案例修正模块：该模块实现对深层判断模块的疑难案例的修正学习功能，对典型案例库中的不适当案例予以修正，对典型案例库中没有的案例予以补充。

诊断结果或建议输出模块：该模块将诊断结论信息或建议返回给用户。

系统后台维护模块：维护整个系统的后台数据，包括系统注册信息的维护，典型案例一致性的维护、新案例的补充、不适当案例的删除以及有关鱼病资料的图文数据的维护和管理。

15.4　知识库设计

15.4.1　系统知识库

知识库是知识系统的核心，在讨论如何利用关系数据库来表示和存储系统知识之前，有必要对基于案例的鱼病系统知识库结构做一下分析。在传统的专家系统理论中一般把知识库和数据库区分开来，认为知识是知识，数据只是数据。在基于案例的系统的研究中，案例是知识

的一种模式，其直观地表现为数据，所以本文将知识库和数据库合二为一，形成综合知识库，也称系统知识库。[3]

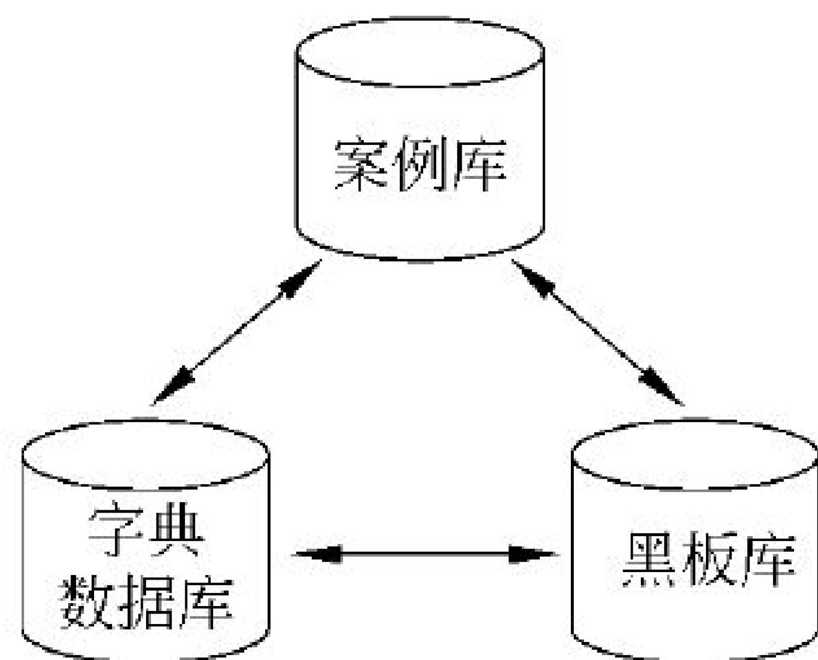


图 15-5　系统知识库

基于案例推理的鱼病诊断专家系统知识库由字典数据库、案例库和黑板库组成，如图 15-5 所示。

(1) 字典数据库主要存放有关鱼病诊断问题领域要素的事实性数据，包括：鱼病症状，疾病，病因，防治方法以及其他相关事实数据。

(2) 案例库中存储大量的典型案例，一个典型的 CASE 一般包含三方面的信息：①问题的说明信息，即问题的初始条件；②问题求解的目标；③达到该目标的解决方案。

(3) 黑板库也叫动态库，它是反映当前问题求解状态的集合，用于临时存放求解问题所需的各种初始数据或证据，以及求解期间由系统产生的各种中间信息，它是推理机选用知识的依据，也是解释机制获得推理路径的来源。

系统知识库的三个组成部分彼此联系，构成一个闭环系统。案例数据借助于字典数据库中的事实性数据对案例进行描述，黑板库中的临时数据也依赖于字典数据。反过来，字典数据库中的数据结构发生变化时案例库和黑板库中的数据也会跟着改变。在问题求解过程中，案例库和黑板库之间会发生数据的读取、交换和修改，以完成目标问题的求解和案例库的更新维护。[4~6]

分析了系统知识库的结构和组成，下面将在关系数据库的基础上来建造知识库，充分利用关系数据库管理系统功能，可以方便知识库管理系统的设计。

15.4.2　数据库概念结构设计

鱼病诊断专家系统既是一个诊断推理的专家系统，也是一个数据库应用系统。但是其诊断推理过程比一般信息系统的应用处理过程要复杂得多。因此，基于案例推理的鱼病诊断专家系统数据库设计过程更加复杂，其设计过程中既要考虑鱼病诊断问题的复杂性，还要考虑系统案例修复功能中新的问题案例一旦确诊后如何转化为可用案例的情况。

利用数据库技术建立鱼病专家系统知识库，是通过数据库本身的技术，如关联、过滤机制和索引技术等实现简单的知识推理。另外，可以方便地对这些知识进行增加、删除、修改、浏览等操作，实现专家系统知识库的维护，极大地简化系统设计和维护人员对已有知识的访问过程和管理、降低了维护难度。[7]

15.4.3　数据库逻辑结构设计

通过对鱼病系统领域知识的分析，设计了 8 张事实表(对应系统字典数据库)和 5 张案例表(主要用于案例存储，对应系统案例库)，2 张临时表(对应系统黑板库)。表中带下划线的属性字段是主键(见图 15-6)。

15.4.4　数据表之间的依赖关系

通过上述数据逻辑结构的设计分析，可以得出以下如图 15-6 所示的数据表之间的依赖关系。



图 15-6　数据表之间的依赖关系图

15.5 系统实现

以下是所实现系统的几个界面图示例(见图 15-7～图 15-9)。



图 15-7　鱼病案例基本信息输入页面



图 15-8　鱼病案例典型症状输入页面





图 15-9　鱼病案例推理结果输出页面

15.6　小结

根据专家系统开发的技术要求，以 Jbuilder9＋Dreamweaver MX＋SQL 2000 为开发工具，在 Windows2003 Server＋Apache＋Tomcat5.0 系统/客户开发平台下，采用 JSP＋JavaBeans 的应用层模式，建立了基于 Web 的 B/S/S 三层体系结构的鱼病诊断专家系统。

参考文献

1. 温继文. 基于知识的鱼病诊断推理[D]. 中国农业大学，2003
2. 马洪兵，张秋玲. HTML 语言与 Web 站点开发技术[M]. 北京：清华大学出版社，1998
3. 杨善林，倪志伟. 机器学习与智能决策支持系统[M]. 北京：科学出版社，2004
4. 张本生，于永利. CBR 系统案例搜索中的混合相似性度量方法[J]. 系统工程理论与实践，2002，3：131～136
5. 章曙光，钱权，方瑾等. 范例推理中基于时序的范例匹配方法模型[J]. 小型微型计算机系统，2003，4：640～642
6. 钟诗胜，王知行，黄炽强等. 基于模糊相似优先的实例检索模型[J]. 计算机研究与发展，1998，9：56～59
7. 萨师煊，王珊. 数据库系统概论[M]. 北京：高等教育出版社，2000

第16章

基于呼叫中心的鱼病智能诊断系统

本章将对基于呼叫中心的鱼病诊断系统的结构、功能模块、工作流程和知识库进行具体设计，配置系统硬件和软件平台，核心是各主要功能模块的程序实现，最后是系统的实现。

16.1　系统结构设计

面向呼叫中心的鱼病智能诊断系统按系统的结构层次划分为：用户层、传输层、呼叫中心层、鱼病智能诊断层和服务器层。如图16-1所示，下面介绍各结构层次。

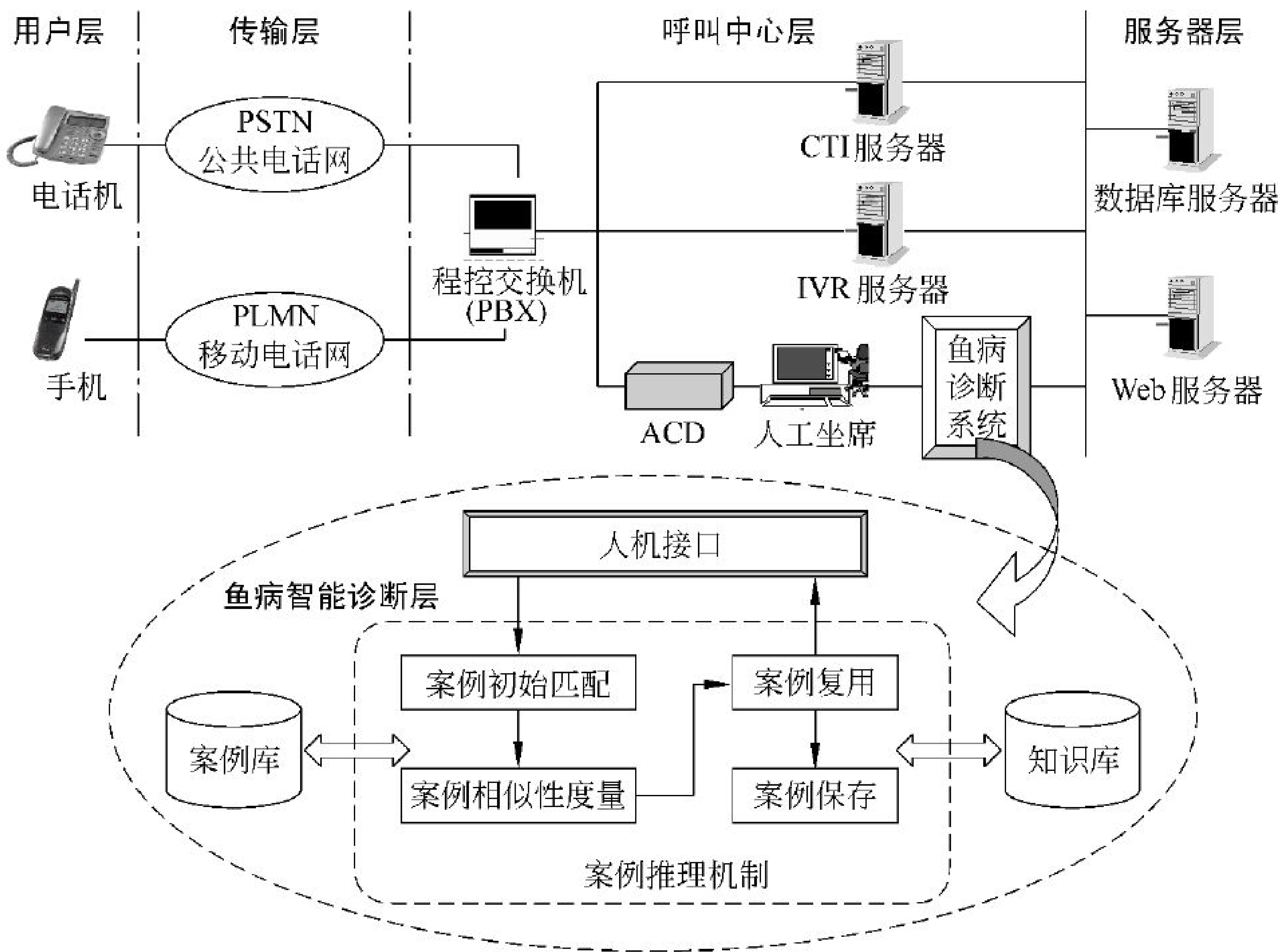


图16-1　系统结构

（1）呼叫中心层。本层由六部分组成：程控交换机（PBX）、自动呼叫分配器（ACD）、交互式语音应答（IVR）服务器、计算机语音集成（CTI）服务器、人工坐席（Agent）和原有系统主机。[1~3]

（2）鱼病智能诊断层。这层是本系统的核心部分，其推理机部分由鱼病症状特征辨识、案例初始匹配、案例相似性度量、案例复用和案例保存组成。该层还包括人机接口、鱼病案例库和知识库。



(3) 服务器层。该层包括数据库服务器和 Web 服务器,其中数据库服务器完成对系统数据库、知识库和案例库的访问和维护,Web 服务器完成网络系统的运行和维护。

(4) 用户层。访问呼叫中心的用户是养鱼户,他们可以通过固定电话或手机访问本系统。

(5) 传输层。本层是指养鱼户与本系统进行信息交互要经过的通信网络,其中使用固定电话的养鱼户通过公共电话网(PSTN)访问本系统,使用手机的养鱼户通过移动电话网(PLMN)访问本系统。

16.2 功能模块设计

按照系统各个组成的功能划分,包括呼叫中心人工服务模块、鱼病诊断信息自动语音查询模块、鱼病诊断症状信息输入模块、鱼病智能诊断推理模块、诊断结果信息输出模块、数据库管理与维护模块。如图 16-2 所示。

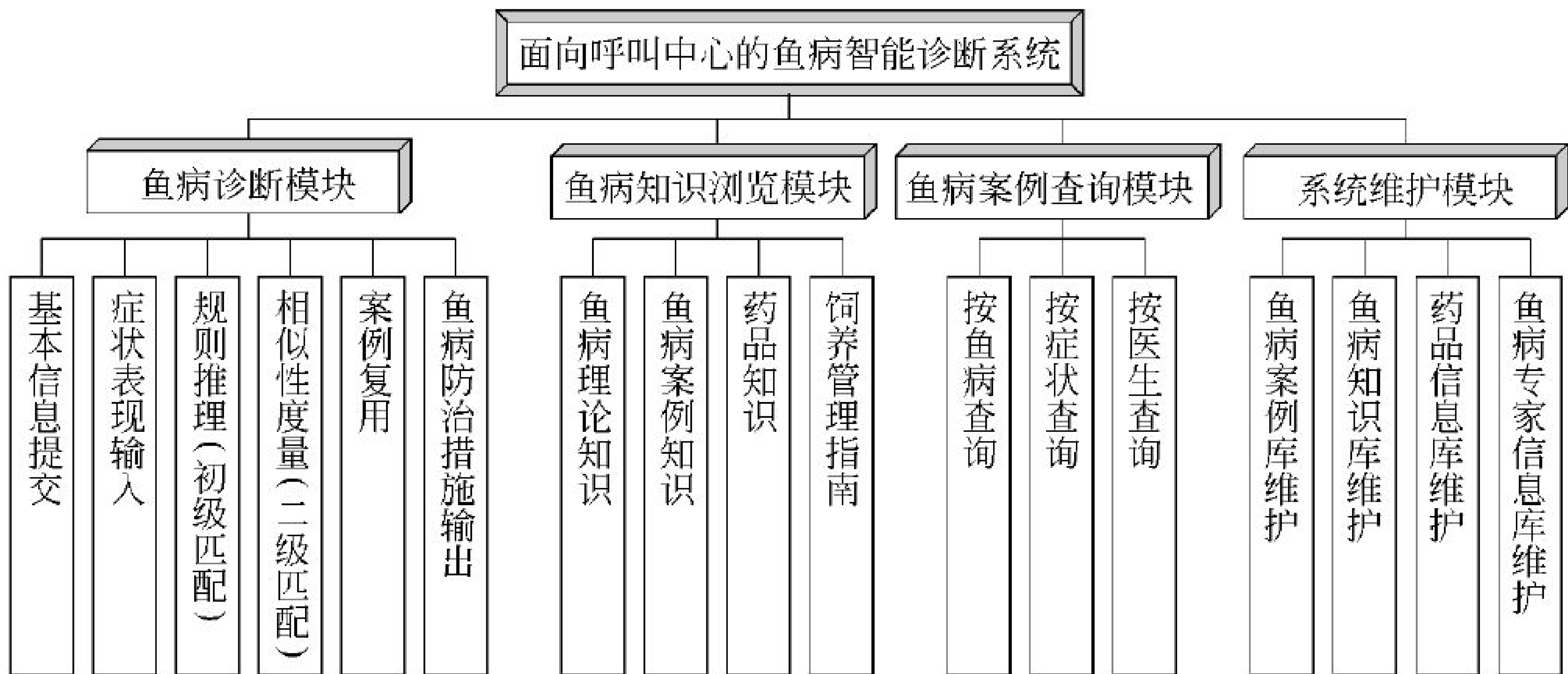


图 16-2 系统功能模块

16.2.1 鱼病诊断模块

鱼病诊断该模块包括鱼病基本信息提交、症状表现输入、规则推理(初级匹配)、相似性度量(二级匹配)、案例复用和鱼病防治措施输出六部分。养鱼户提交的鱼病症状描述信息经过坐席人员的鱼病症状特征辨识,从自然语言中提取出规范的鱼病症状特征信息,再由坐席人员将其输入系统。经过鱼病智能诊断推理模块的推理,新案例在案例库中找到最相似的源案例,并复用源案例的结果和防治措施,该模块将诊断结果信息和防治返回到系统界面上,再由坐席人员反馈给养鱼户。该功能模块根据两级索引来运行,首先将新案例的症状特征在第一级索引的规则集合中进行归类,由此得出和目标案例相似的一类源案例群,初步确定疾病,实质上这是正向推理;再根据初步确定的疾病,在案例库中根据分类标识进行匹配,若完全匹配,确诊疾病,并给出相应的治疗方案,若不完全匹配,则根据分类标识用最近邻法计算找出和目标疾病相似度最大的案例,并进行适当修正,重新进行检索,该检索在大幅度缩小的搜索空间里进行,检索效率明显提高。这一步是反向推理,在案例库中来证实上一步得出的结论是否正确。[4]

推理算法如下：

步骤1：输入新案例，按照索引生成的有关规则生成新案例的规范化描述。

步骤2：根据症状部位判断得出属于哪一个案例组：若有1个，转步骤3；若多于1个或没有（$i=1$，转步骤4；$i=2$，转步骤6）。

步骤3：根据具体症状匹配计算得出最相似案例：若有1个，转步骤5；若多于1个或没有（$i=0$，转步骤4；$i>1$，转步骤6）。

步骤4：向用户请求更多症状信息，转步骤1。

步骤5：输出最相似案例信息。

步骤6：进行基于知识的检索。

步骤7：输出相似案例组和知识检索结果。

步骤8：用户反馈。

16.2.2 鱼病知识浏览模块

鱼病知识浏览模块包括鱼病理论知识、鱼病案例知识、药品知识和饲养管理指南等，养鱼户可以获得这四个方面的知识。其中，鱼病理论知识中包含着常见多发的鱼病的病名、病因、症状表现、流行情况、防治措施和用药方法；鱼病案例知识中包含着本系统所有的案例知识；药品知识中包含着常见药品的名称、功能特性和使用方法；饲养管理指南中包含了水质调控、饲养管理、营养与饲料等内容，可以指导养鱼户的日常饲养管理。

16.2.3 鱼病案例查询模块

鱼病案例查询模块实现对系统案例知识库的查询功能，用户可以通过鱼病、症状表现和诊断医生三个方面对案例进行查询。

16.2.4 系统维护模块

系统维护模块实现对鱼病案例库、知识库、药品信息库和专家信息库的维护，包括数据的添加、修改与删除等操作。

16.3 工作流程设计

本系统的工作流程如下：

（1）养鱼户拨叫呼叫中心号码；

（2）呼叫信号经过程控交换机（PBX）进入呼叫中心；

（3）呼叫信号在自动呼叫分配器（ACD）内进行排队，依次转接给坐席人员；

（4）坐席人员根据用户需求，进行鱼病智能诊断系统的操作。若进行鱼病诊断，则转（5）；若进行鱼病知识浏览，则转（11）；若进行鱼病案例知识查询，则转（12）；

（5）坐席人员通过系统主界面输入案例基本信息；

（6）坐席人员获取来自养鱼户的鱼病症状信息，进行特征辨识，并提交到系统；

（7）进行初始匹配（规则推理），若有结果，则转（8）；否则，返回（6），请求更多信息；

（8）进行相似性度量，若有结果，则进行（9）；否则，则进行（8）；否则，返回（6），向养鱼户请求更多信息；



(9) 进行案例修改；
(10) 输出鱼病防治措施；
(11) 选择进入“鱼病理论知识”、“案例知识”、“药品知识”或“饲养管理指南”；
(12) 选择按照“鱼病”、“症状”或“医生”进行查询。
本系统的工作流程示意图如图 16-3 所示。

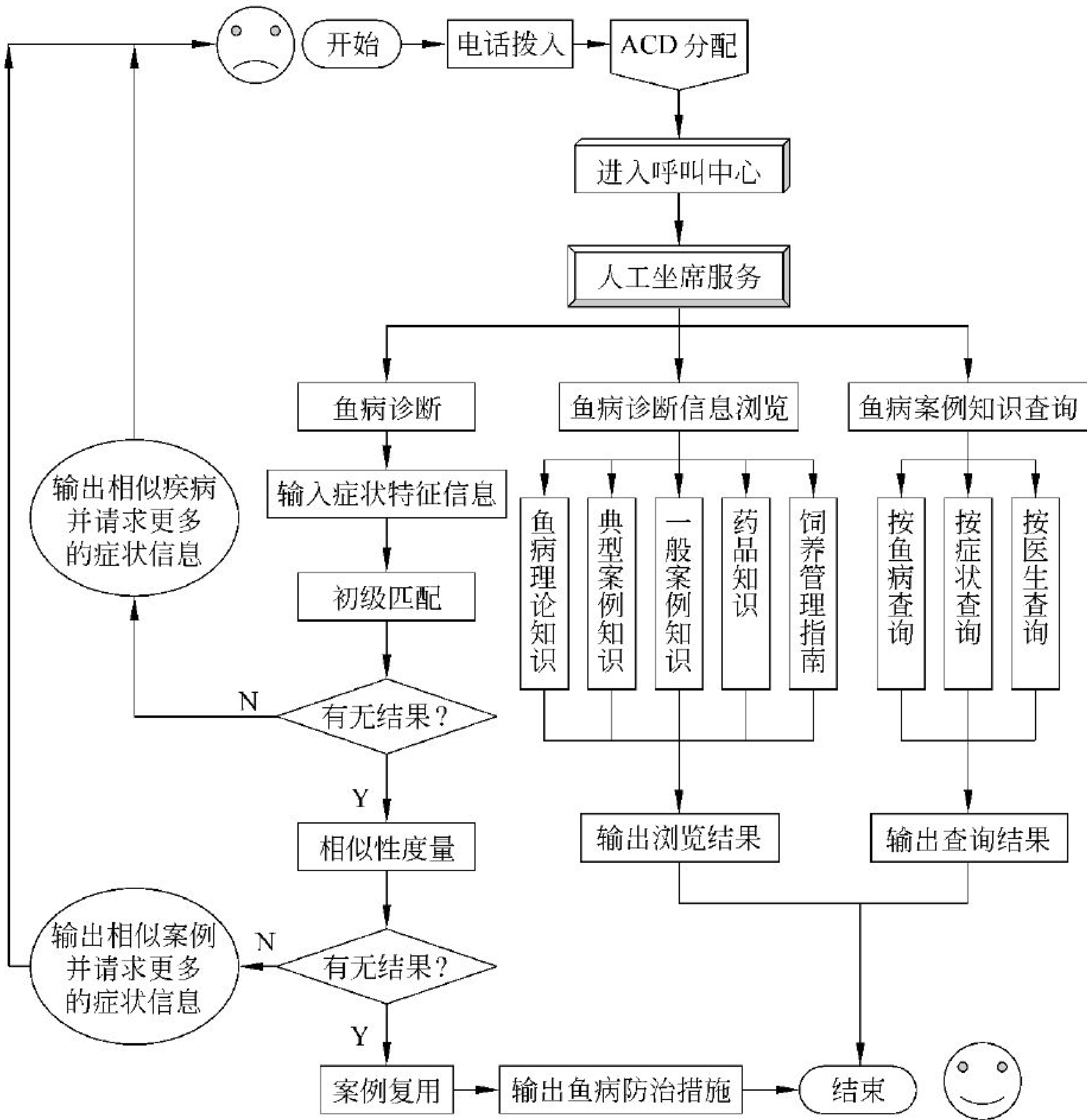


图 16-3　系统工作流程

16.4　知识库设计

鱼病诊断专家系统既是一个诊断推理的专家系统，也是一个数据库应用系统。但是其诊断推理过程比一般信息系统的应用处理过程要复杂得多。因此，基于案例推理的鱼病诊断专家系统数据库设计过程更加复杂，其设计过程中既要考虑鱼病诊断问题的复杂性，还要考虑系统案例修复功能中新的问题案例一旦确诊后如何转化为可用案例的情况。

利用数据库技术建立鱼病专家系统知识库，是通过数据库本身的技术，如关联、过滤机制和索引技术等实现简单的知识推理。另外，可以方便地对这些知识进行增加、删除、修改、浏览等操作，实现通过数据库技术的专家系统知识库的维护，极大地简化了系统设计和维护人员对已有知识的访问和管理、维护难度。该实现方式从某种意义上讲实现了通过数据库技术对知识库本身的维护。

由鱼病诊断的知识表示模型可知，鱼病案例是对一个鱼病事件(问题)的描述和表示，一个典型的鱼病案例 C 可用一个七元组 CASE＝(DO,M,W,D,S,P,T)来描述。其中，DO 表示对象信息，即什么品种和规格的鱼；M 表示时间信息，即什么季节；W 表示水环境信息，由pH、溶氧度等指标构成；D 表示疾病信息；S 表示症状信息；P 表示病因信息；T 表示治疗信息。其中最主要的元素是：诊断对象，症状，疾病，病因，治疗方案。

(1) 初始鱼病案例表(案例 ID，诊断时间，鱼塘地点，养鱼户，病鱼种类，病鱼规格，水质信息，症状表现，疾病，病因，治疗方法，医生)，即 case_base(case_id，time，place，farmer，variety，weight，water，symptom，disease，pathogeny，therapy，expert)。

(2) 鱼病 X 案例判别表(案例 ID，症状 1，症状 2，…，症状 n)，即 case_base_x(case_id，symptom1，symptom2，…，symptom$n$)。这里每种鱼病都有一个案例表。

(3) 鱼病诊断知识表集(鱼病 ID，鱼病，发病时期，易感鱼种，症状表现，病因，治疗方法，用药)，即 disease_knowledge(disease_id，disease，time，variety，symptom，pathogeny，therapy，medicine)。

(4) 药品知识表(药品 ID，药品名称，功能，使用方法)，即 medicine_knowledge(medicine_id，medicine，function，method)。

(5) 鱼病专家信息表(专家 ID，姓名，专业特长，联系方式)，即 expert(expert_id，name，capacity，telephone)。

(6) 养鱼户信息表(养鱼户 ID，姓名，地点，养殖种类，养殖规模，联系方式)，即 farmer(farmer_id，name，address，variety，scale，telephone)。

16.5 系统实现

16.5.1 系统开发环境和开发技术

本系统在 Windows XP＋IIS＋.NET Framework＋MS Access 2003 环境下进行开发、编译及调试。

本系统开发采用了以下技术。

(1) ASP.NET——本系统的功能层采用 ASP.NET 技术实现。ASP.NET 不是 ASP 的升级版本，它为 Web 开发创造了一种全新的、更加灵活的基础结构。ASP.NET 是编译执行的 Web 服务器端开发工具，当用户访问 ASP.NET 文件时，执行编译后的文件，而不是源文件。

(2) ADO.NET——本系统的数据层采用 ADO.NET 技术实现。ADO.NET 是基于 ADO 的数据访问技术，它利用 Microsoft.NET Framework 的托管代码创建，是一个优于 ADO 的更为稳健的数据库访问平台。ADO.NET 的设计方式使它能够解决松散耦合的 Web 应用程序环境中的问题，在这种环境下，不同组件在本质上是非连接的。

16.5.2 呼叫中心系统的环境配置

呼叫中心系统的开发采用了青牛软件公司提供的 USE 呼叫中心为中间件，青牛 USE 是一个面向应用开发商和系统集成商的呼叫中心中间件平台产品。青牛 USE 为应用开发提供了强大的底层支持，包括实时服务路由(语音话路控制、文本 Chat、同步浏览等)、异步服务路



由(邮件、Voice Mail、SMS 等)、实时监控和系统运行统计等。

本系统的局域网实施采取线形拓扑结构,一根主线上连接着各种相应的设备。系统网络架构中的各种服务器包括数据库服务器、应用服务器用以提供整个系统的数据和文件。管理台负责对硬件设备和软件环境进行监督和维护。电话用户通过公用电话网、移动用户通过移动通信网对呼叫中心进行访问,其访问要通过模拟交换机连接 CTI 服务器。

16.5.3　系统运行界面

本节给出一些系统运行界面的示例(见图 16-4～图 16-6)。



图 16-4　呼叫中心坐席界面



相似性度量结果

case_id	time	place	farmer	variety	weight	symptom	disease	pathogy	therpy	expert	posibility
28	2002-4-25		边德超	鲢鱼	鱼苗	厌食，有死亡，腹部肿胀，肛门红肿，肛门有黄色液体流出，肠道充血，肠道无食物，肠道有脓液	肠炎	摄食量太大，水质未处理，使料台及整池水质恶化	内服25%大蒜素2‰3-5天，外用二氧化氯	赵立明	0.1538556
29	2002-5-1		马铁锁	白鲢	300g	池边打转，有死亡，肠道积水，肠壁充血，肠道有黄色黏液	肠炎	氨氮中毒，水质不良，水中缺少白鲢饵料	泼洒降氨氮药物结合土霉素外用	陈成勋	0.1818314
30	2002-5-2		边德超	草鱼	70g	厌食，死亡，肛门红肿，肛门外突，肠道积水，肠道有脓液	肠炎	饲料中毒	外用消毒剂；内服呋喃，氟派酸，大蒜素	刑克智	0.1904907
31	2002-4-5		经立臣	鲤鱼	60g	厌食，肛门红肿，腹部充血，肠道充血，肠道积液，肠道无食物	肠炎	连天阴雨，饲料霉变	连服呋喃，氟派酸1%5天，漂白精0.7ppm全泼一次	孙世娟	0.16
32	2002-6-9		张奇林	鲢鱼	250g	厌食，死亡，离群独游，肠壁充血，肠内无食物，肠黏液多，肠有脓	肠炎	连天阴雨造成饵料霉变后投喂诱发	内服呋喃唑铜，大蒜素，氟派酸5-7天，外用二氧化氯全泼	赵立明	0.1481569
33	2002-6-2		沈宝成	鲤鱼	300g	厌食，死亡，肠道发红	肠炎	连天阴雨，饲料霉变	痢特灵：1斤大蒜素+2斤氟派酸/3千斤料	陈成勋	0.2353163
34	2002-5-14		马铁锁	鲤鱼	120g	厌食，肠道发红，肠内有剩料	肠炎	连天阴雨造成饵料霉变后投喂诱发	敌百虫+亚铁	孙世娟	0.1818182

图 16-5　相似性度量结果

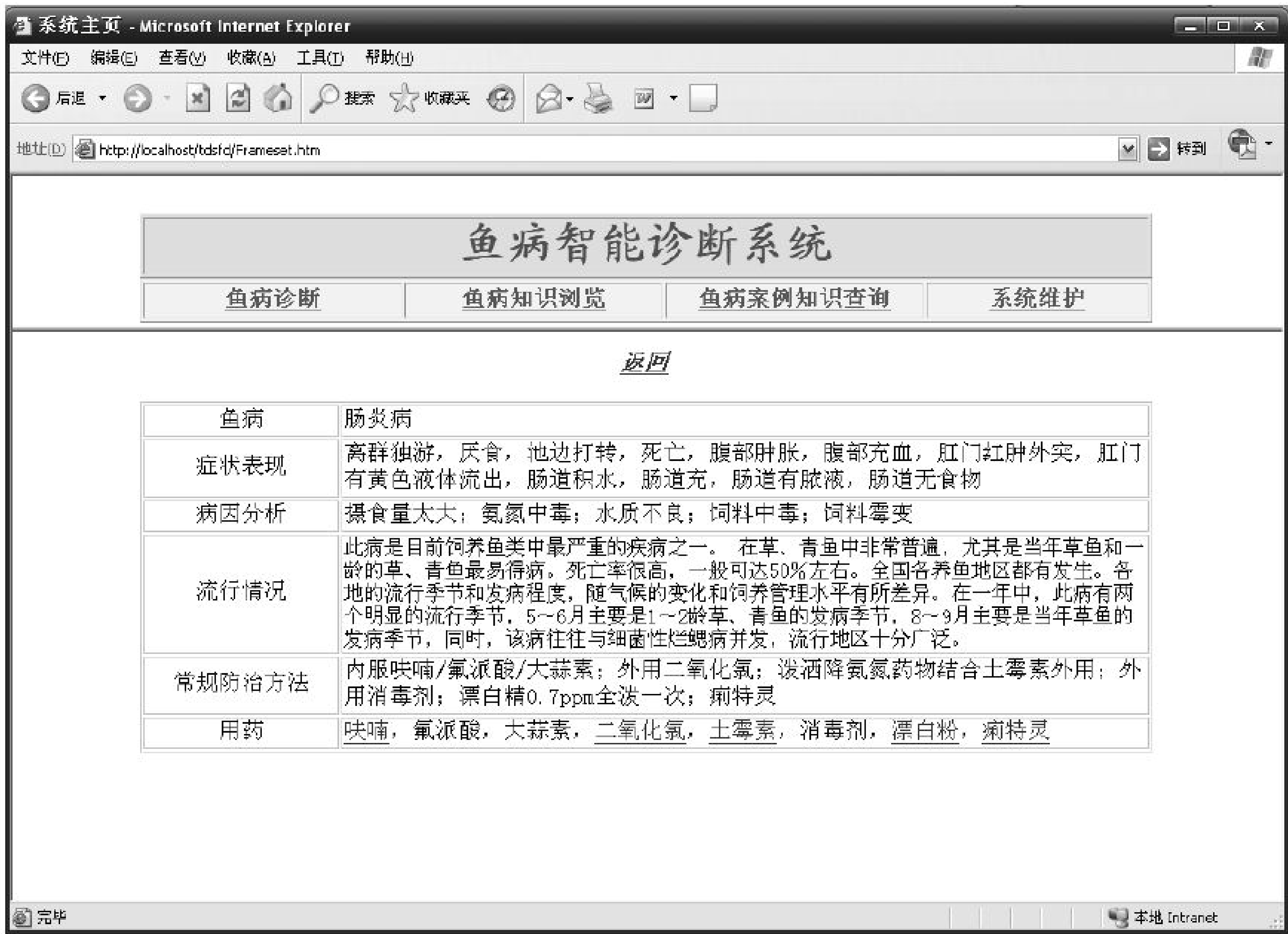


图 16-6　鱼病信息浏览窗体

16.6　小结

本章对基于呼叫中心的鱼病诊断系统的结构、功能、工作流程进行了具体设计，配置了系统硬件和软件环境。采用 ASP. NET、ADO. NET 技术，以 Windows XP＋. NET Framework＋MS Access 2003 为开发平台，开发了面向呼叫中心的鱼病智能诊断系统，实现了预期功能。

参考文献

1. 许娓. 呼叫中心的核心技术及组成[J]. 电脑与信息技术，2001，6：15～18
2. 宋曙光，毛新军，谭庆平. 传统呼叫中心与 Internet 呼叫中心的一体化[J]. 现代电信技术，2000，6：1～5
3. 耿玉亮，李伟生. 基于 Internet 平台的呼叫中心的研究[J]. 计算机应用研究，2003，1：55～57
4. 唐智勇，韦东鹏，王红胜. 基于线索和改进最相邻近法的案例检索[J]. 计算机工程与应用，2000，4：78～80
5. 陈敏，阳国贵，吴泉源. 基于事例推理中事例组织与管理的研究[J]. 计算机工程，1994，5：56～62
6. 孙俊杰，王新允. 基于 Web 的呼叫中心的功能与实现[J]. 中国数据通信，2001，5：50～55
7. 张健，傅泽田，张小栓. 农业病虫害远程呼叫与咨询系统的设计与分析[J]. 计算机应用研究，2004，11：181～183

第17章

移动式鱼病诊断专家系统

本章针对目前水产养殖专家系统存在的问题，结合用户实际需求，研究了移动式鱼病诊断专家系统的知识获取和推理机制的理论方法；提出移动式鱼病诊断专家系统的总体结构方案，确定其软硬件体系结构和实现方案；设计并实现了该专家系统。

17.1 移动式鱼病诊断系统需求分析

17.1.1 需求特点

根据实地考察、走访用户和参考有关文献，结合前期相关示范项目的实施经历和经验[1]，总结出鱼病专家系统用户需求有如下主要特点：

(1) 养殖场和养殖户多地处偏远且分散、隔离，交通不便，极其缺乏鱼病专家和技术人员，因此迫切需要鱼病专家系统的援助，同时也决定了该系统的使用方式是分布的；

(2) 养殖地一般远离鱼病诊治中心，但同时许多病害的发生需要现场及时诊治，所以，能满足随时随地进行鱼病诊断的便携式专家系统受到用户青睐，这需要在便携设备上嵌入专家系统，或通过便携设备访问后端专家系统；

(3) 伴随网络和通信设施的逐步普及，阻碍农业信息化服务与应用的所谓“农业信息最后一公里”问题也行将得到解决，可以说，网络化是构造分布式系统的技术核心，同时也是分布式应用的基础；

(4) 对某些疑难杂症，限于机器智能的诊断能力，还需要人类专家的参与。因此，该专家系统在功能上需要做进一步的分布处理，即提供呼叫中心、专家坐诊等功能。

由此可见，整个专家系统需要提供功能分布、知识分布、推理系统分布、应用分布等机制，才能够满足用户在功能和应用上的需求。

17.1.2 系统逻辑结构

根据上述需求特点，设计了如图17-1所示的系统逻辑结构。由图17-1可以看出，该系统基本实现了各种分布式用户需求：

(1) 通过知识获取系统获得系统的公共知识，存入公共知识库。这是整个系统的关键，是实现其他子系统和知识分布、推理分布的基础，是本章的研究重点。

(2) 每个子系统通过不同机制获得适合自身需要的知识和数据信息，并存储在自身的知识库、数据库中。例如，移动专家子系统是适合于移动用户的专家系统平台，它在运行环境、数据格式、用户界面等方面不同于固定式专家系统，因此必须有自己的知识库、数据库和推理机等。

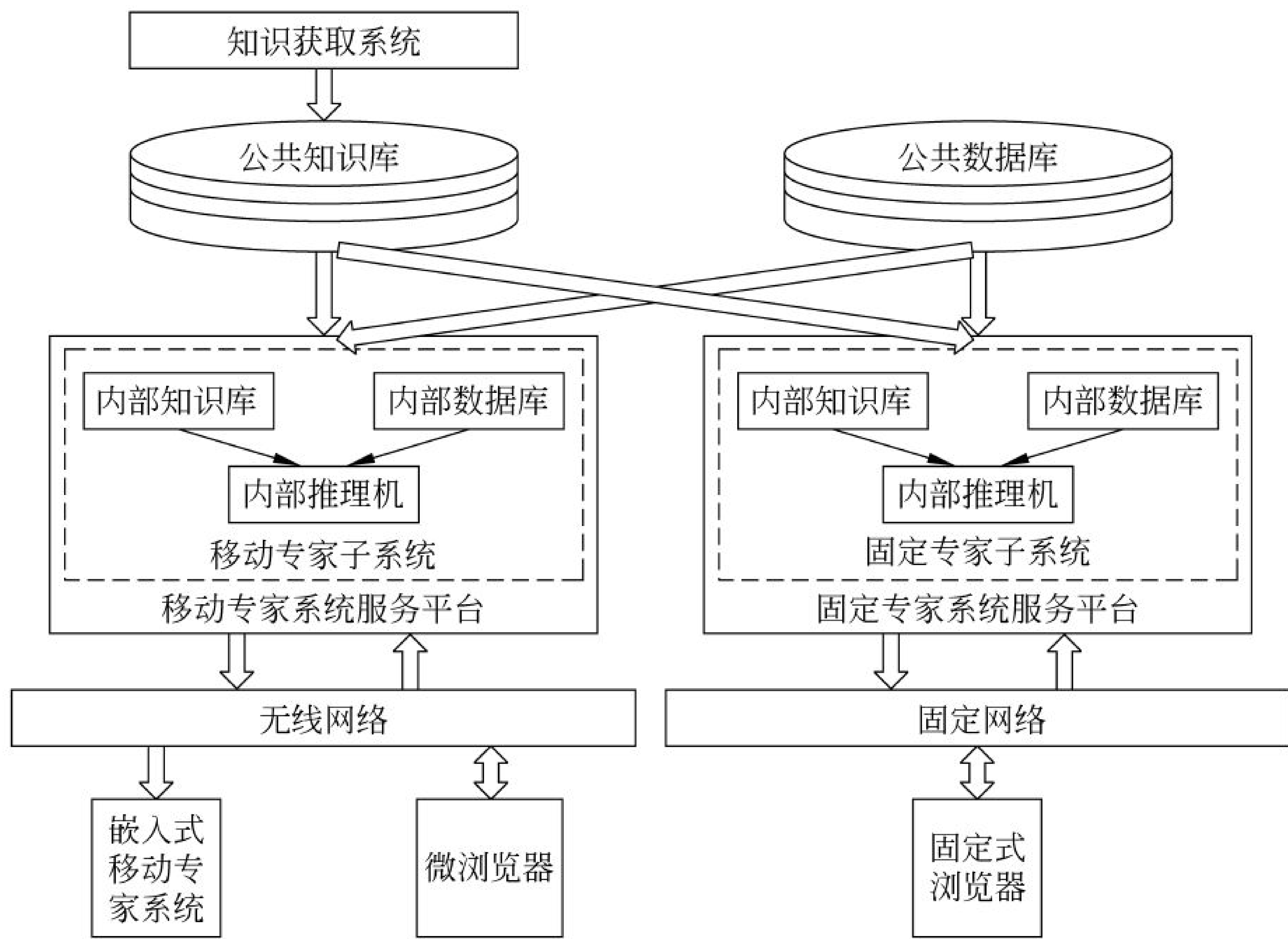


图 17-1　系统逻辑结构

（3）系统根据不同的应用环境（如网络环境）和应用要求，可提供三种不同的应用模式：本地移动应用模式、移动 B/S 或 C/S 模式、固定 B/S 模式，从而实现了应用分布的需求。

（4）嵌入式移动专家系统是一个完整的本机系统，可以通过无线网络下载、更新，支持随时随地的现场诊断，不受网络环境的影响，实现了功能分布。

17.1.3　技术难点与实现策略

（1）由于嵌入式专家系统运行在资源有限的手持计算设备（如智能手机、PDA 等）上，无法承受大规模的复杂计算和存储大容量数据，因此不宜完全采用传统专家系统的推理机和知识库结构。本系统充分利用服务器端的计算能力，首先通过服务器上的神经网络训练学习获取知识，以神经网络推理模式存储相关知识，并将该知识库和推理系统移植到移动设备中，不但可以减少数据量，而且可以提高推理效率。[2]

（2）尽管服务器有较高的计算能力，但是由于规模庞大的神经网络训练非常耗费时间，因此，采用粗糙集方法先对数据作约简预处理，从而减小神经网络规模，这样不但可以大大提高系统效率，而且还可以提高专家系统处理不确定性、不完整信息的能力。

（3）由于系统整体上是一个分布式、跨多平台的架构，因此开发平台的选择是个关键。考虑到网络性能和平台适应性，本系统选择以 Java 平台为核心的开发技术。

17.2　系统整体结构设计

17.2.1　系统物理架构

移动式鱼病诊断专家系统物理架构如图 17-2 所示。它是由移动客户端、固定客户端、网络通信系统（Internet、无线网络）和服务器端等组成的物理分布系统。



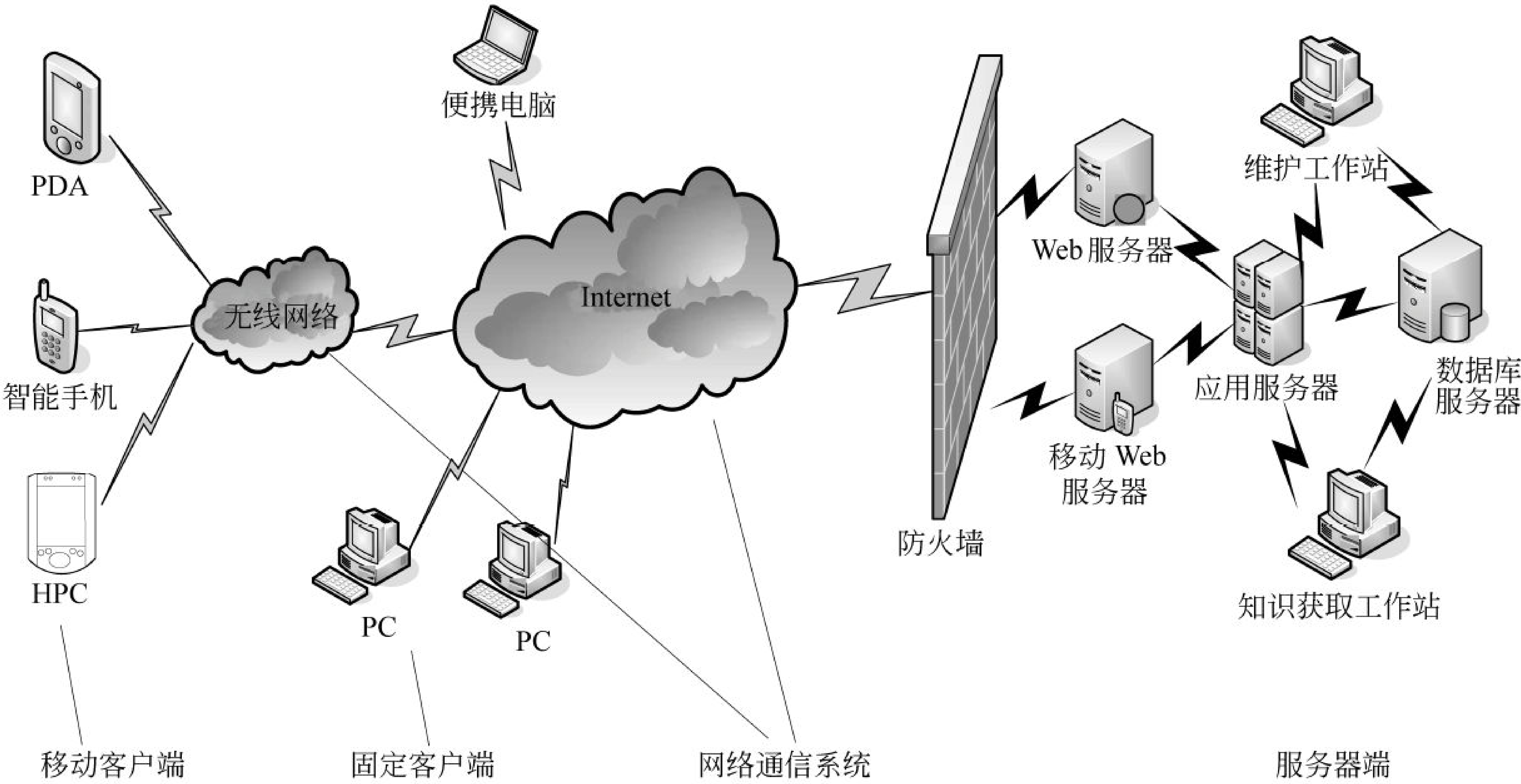


图 17-2　移动式鱼病诊断专家系统物理架构

17.2.2　系统服务模式设计

该系统可以提供 B/S、C/S、本地(本机嵌入式系统)三种访问移动专家系统的模式,如图 17-3 所示。

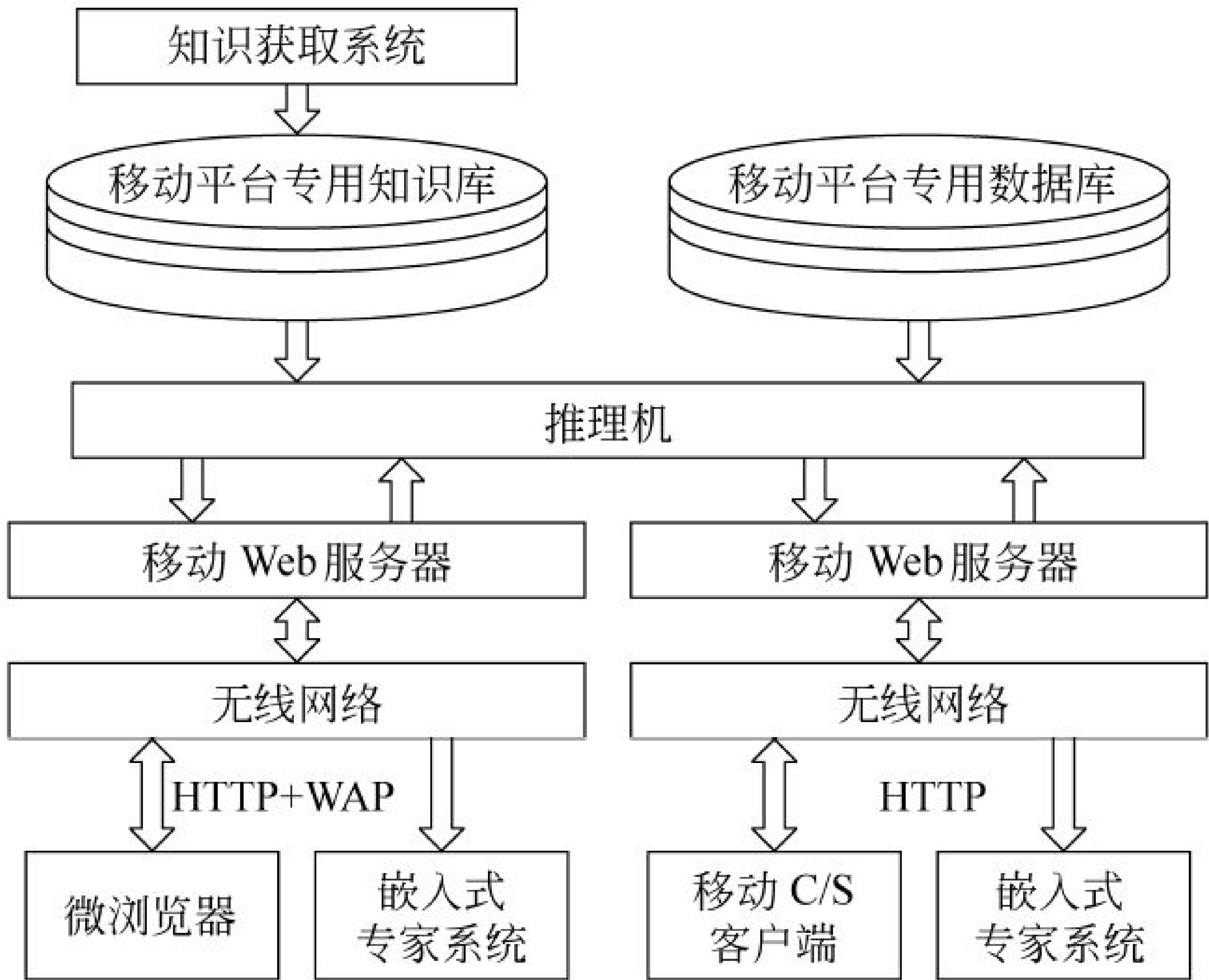


图 17-3　系统服务模式

移动式专家系统由服务器端子系统、客户端子系统和网络通信服务系统组成。可以满足用户在有稳定的无线网络连接环境和无稳定的无线网络连接环境下现场诊断的需求。其中服务器端子系统、客户端子系统是系统实现的关键。[3]

1. 服务器端子系统

服务器端子系统由知识获取系统、服务器端专家系统和移动 Web 服务器组成。

(1) 知识获取系统

鱼病知识获取系统是整个专家系统实现的关键，其原理和设计见上一章。具体实现时，考虑到系统采用了大规模智能计算技术(神经网络、粗糙集、蚁群算法等)，因此在算法实现上采用效率较高的 C/C++ 语言，鱼病案例数据和相关辅助与结果数据以 SQL Server 数据库存储。

(2) 服务器端专家系统

服务器端专家系统由移动专家系统专用知识库、数据库和推理机组成。移动平台专用知识库中的数据由知识获取系统提供和更新，移动平台专用数据库存储着移动专家系统的公用数据(如事实数据、资料数据、中间结果数据等)，存储在 SQL Server 数据库服务器中。其推理机用 Java 中间件实现。

(3) 移动 Web 服务器

移动 Web 服务器提供用户访问的用户界面，以 Apache+Tomcat 作为 Web 服务器，后端以 SQL Server 作为数据库服务器，用户界面采用 JSP 和 Servlet 技术实现。

2. 客户端子系统

客户端子系统包括三种类型的客户端，即嵌入式鱼病诊断专家系统、微浏览器型客户端、移动 C/S 型客户端。

(1) 嵌入式鱼病诊断专家系统

该嵌入式移动专家系统是一个完整的本机专家系统，通过在移动计算设备(如智能手机、PDA 等)中嵌入知识库、数据库和推理机等实现，可以通过无线网络下载、更新。它支持随时随地的现场诊断，使用中不受网络环境的影响。其结构如图 17-4 所示，实现了系统在知识库功能和应用上的移动要求。

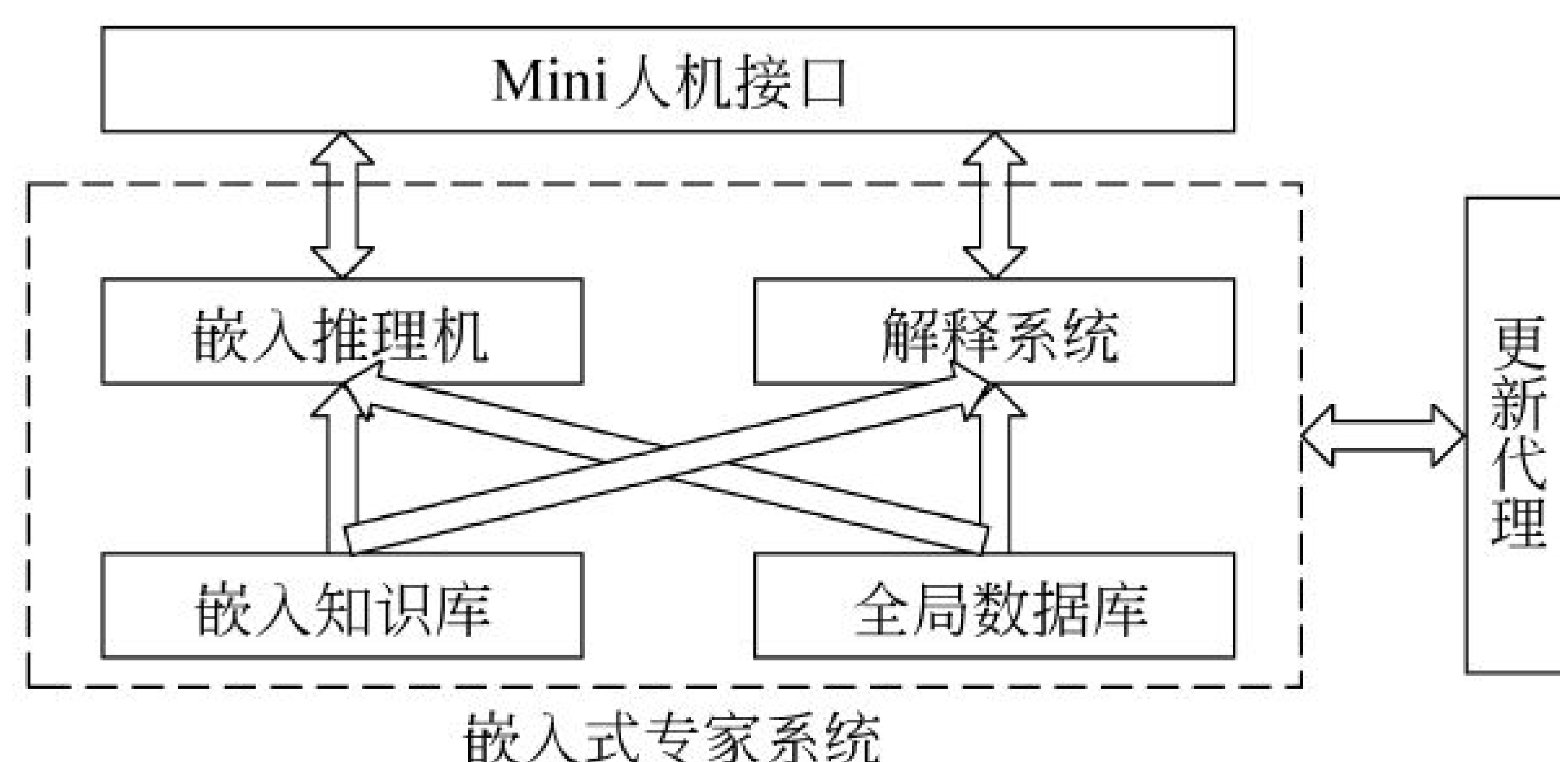


图 17-4　嵌入式鱼病诊断专家系统结构

该系统采用 J2ME 技术实现，具有良好的平台适应性和移植性。其知识库、推理机通过服务器端子系统的知识获取与推理机制产生，因此具有与服务器端相当的专家推理能力。但由于采用了神经网络知识表示和推理机制，无须存储大量的案例数据和推理规则，只需存储神经网络知识库和相关公用数据即可，神经网络推理机算法简单、容易实现，对资源有限的移动计算设备不构成负担。

系统数据存储采用 J2ME 特有的“记录管理系统(RMS)”数据格式和文本数据库格式混合存储方式，既保证了存储效率又满足了数据更新的需要。系统通过移动设备的 Java 应用管理器(JAM)可以通过无线网络下载和更新，保持与服务器端系统版本的一致性，从而不断提高系统性能。[4,5]

(2) 移动 C/S 型客户端

移动 C/S 型客户端应用模式是针对具有无线网络支持但不支持微浏览器的手持移动设



备专门设计的。它实际上包括两个主要的组成部分：移动客户端软件和服务器端专家系统服务。

移动客户端软件，提供用户访问后端服务器专家系统的功能并提供使用界面，用 J2ME 的 MIDlet 小程序实现，只要支持 J2ME(KJava)的手持设备都可以使用；服务器端专家系统服务采用 JSP 和 Servlet 技术实现，它们接收客户端发来的服务请求，收集专家系统推理诊断所需的由用户提供的证据信息，调用专家系统的数据库、知识库和推理机进行推理，并负责向客户端返回推理所得的诊疗结论和解释信息。客户端 MIDlet 程序通过 HTTP 协议访问服务器端 JSP 和 Servlet 程序提供的相关服务，并负责与其交互。

(3) 微浏览器型客户端

微浏览器型客户端应用模式是针对具有无线网络支持并带有微浏览器的手持移动设备设计的。它也包括两部分：移动客户端微浏览器软件和服务器端专家系统服务。

移动客户端微浏览器软件由移动设备本身提供，是设备自带的访问无线因特网的软件工具，其功能和使用方式与普通 Internet 浏览器相近，只是受屏幕尺寸限制，界面功能不如普通 Internet 浏览器，只要设备支持无线网络和 WAP 协议就可以访问无线因特网服务。其服务器端专家系统服务功能与移动 C/S 型客户端应用模式的服务器功能类似，只是使用的协议和服务模式不同而已，即这里使用的是 HTTP 和 WAP 协议，以无线万维网服务为主。微浏览器模式的服务器端实际上是一个 WAP 网站系统，用 JSP、Servlet 和 WML 技术实现，JSP 和 Servlet 服务程序不但负责专家系统的调用，还要为客户端提供界面服务，即以 WML 文件格式将有关信息返回给客户端微浏览器。[7]

17.2.3 系统诊断流程设计

本系统的基本诊断流程为：先进行现场诊断，再做目检，最后是镜检，这是一个逐步深入的诊断过程。

(1) 现场诊断：先做现场调查，获取有关水质、环境、季节、鱼种类等初步信息，根据现场病鱼水中表现，由专家系统作出初步病情判断；

(2) 目检：捞取病鱼，对其外在症状做肉眼识别(必要时可作简单解剖处理)，获取症状信息，输入专家系统，由专家系统给出诊断信息；

(3) 镜检：如果有条件和必要(通过以上步骤无法获得满意诊断)，可以进一步做镜检，通过显微镜等手段做病理检查，根据镜检信息由专家系统作出诊断，并给出结论。

如果通过上述过程不能得出满意和符合实际的诊断，可以使用“呼叫专家”功能，直接与服务器端的呼叫中心系统(不属于本章研究内容，这里只提供功能接口)联系，通过语音方式向坐席专家求助。

此外，借助“远程诊断”功能，还可以与在线专家通过“白板”方式实时交流，以求得帮助。这实际上是一个无线即时消息系统。

17.3 知识获取与推理系统设计

移动式鱼病诊断专家系统中，知识获取和推理机制设计是该系统实现的关键。利用粗糙集理论中的有关方法，首先对神经网络的输入向量进行约简，由此减小整个神经网络结构的规

模;然后据此建立网络结构及其与输入向量和输出向量的对应关系(网络标识),再利用该神经网络从训练数据中获取知识并将信息存储为网络连接权信息(网络权值矩阵);最后利用网络权值矩阵、网络结构和网络标识构造推理模型,形成推理机系统。[6] 如图 17-5 所示。

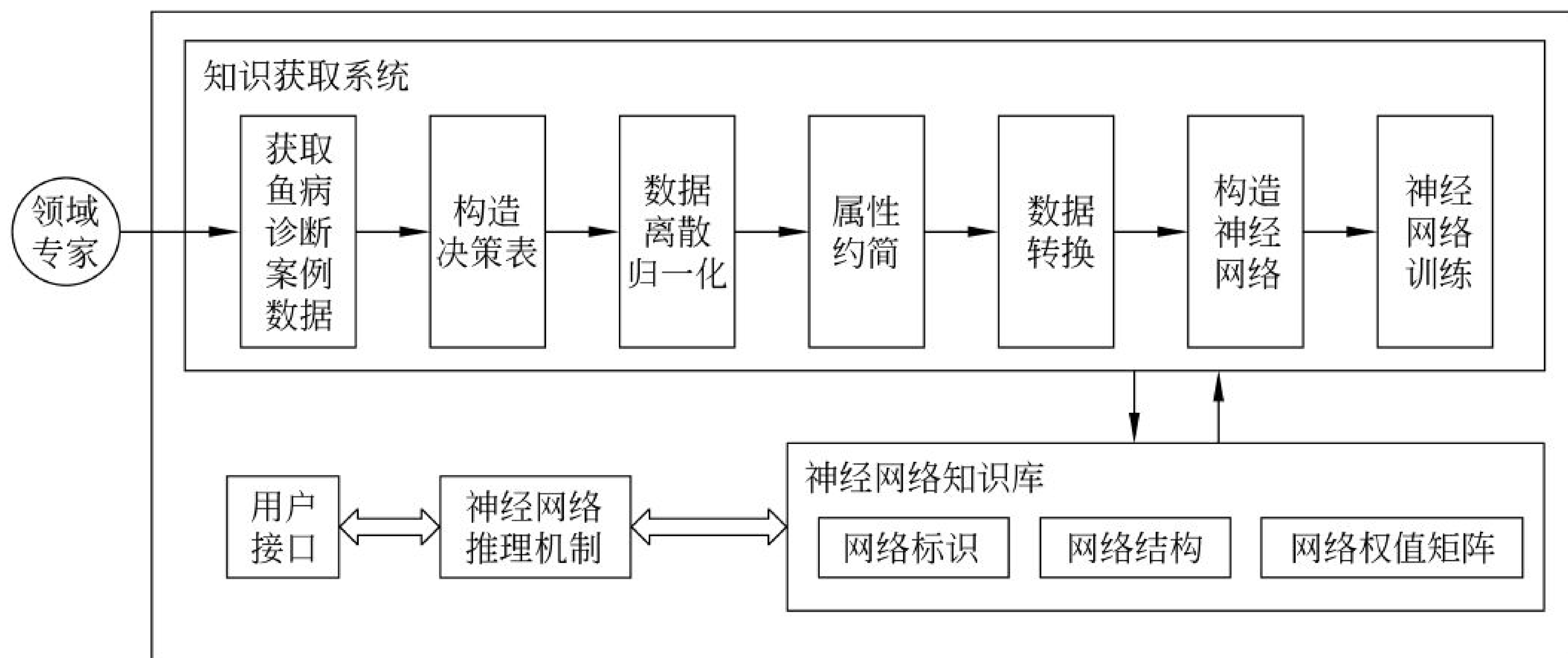


图 17-5　粗糙集-神经网络知识获取与推理系统

系统的主要组成部分如下:

(1) 知识获取模块,由粗糙集预处理机制和神经网络构造与学习模块组成。

(2) 神经网络知识库,由网络权值矩阵、网络结构和网络标识等知识库组成,它们随着训练学习是动态可变的;网络权值矩阵由学习程序产生,是产生推理规则的基本依据;网络结构包括网络参数、数据预处理描述等;网络标识包括推理机制使用的信息对集合,即网络中每一输入输出变量名与实际数据的对应关系。

(3) 神经网络推理机制,由用户证据提取模块、网络知识库数据提取模块和推理程序组成。

(4) 用户接口,即人机对话界面,为系统推理提供所需证据信息和向用户反馈推理结论和解释信息等。

17.3.1　鱼病系统基础数据

案例主要是针对淡水养殖鱼中 46 种常见病、多发病构造的,经验表明养殖户日常所见鱼病绝大多数属于这些疾病。案例库数据经预处理,其结构以及对应属性取值如表 17-1 所示,主要由季节、水温、病鱼规格、病鱼症状和诊断结论(所患疾病)等项数据组成,其中病鱼症状有 84 个,按不同部位分为 8 类:体表(A××)、头部(B××)、鳃部(C××)、腹部(D××)、鳞片(E××)、鱼鳍(F××)、肌肉(G××)、内脏(H××),如“肠道充血发炎”对应内脏部位的症状,其编号为 H02。疾病对应编号为 D××,如 D25 代表“鳃霉病”。

对表 17-1 中的数据,根据专家经验,按“症状接近”原则进行分组,即将 46 种常见鱼病分为 8 组对案例库进行划分,形成缩减的案例空间,如将诊断结果为{出血病(D40)、肠炎病(D35)、白皮病(D31)、赤皮病(D36)、烂鳃病(D28)}的案例分为一组,形成含 127 个案例的决策表,其中条件属性 93 个,决策属性 5 个,分别对应该 5 种疾病。数据经预处理后(见表 17-2)作为粗糙集和神经网络学习的案例数据。



表 17-1　预处理后的鱼病案例属性表

类　别	属　性	描　述	原始值域	离散化值域
季节	S1	春季	“春”	{0,1}
	S2	夏季	“夏”	{0,1}
	S3	秋季	“秋”	{0,1}
	S4	冬季	“冬”	{0,1}
病鱼规格	T1	鱼种	“鱼种”	{0,1}
	T2	鱼苗	“鱼苗”	{0,1}
	T3	成鱼	“成鱼”	{0,1}
	T4	亲鱼	“亲鱼”	{0,1}
水温	Temp	水温	0℃～24℃	{0.0,0.1,…,0.24}
症状	A01～H14	病鱼症状代码	“有”、“无”	{0,1}
疾病	D01～D46	疾病名称代码	“有”、“无”	{0,1}

17.3.2　知识获取

知识获取包括两个主要过程：①基于粗糙集的数据预处理过程；②基于神经网络的学习训练过程，同时产生神经网络知识库。

1. 基于粗糙集的数据预处理

基于粗糙集的数据预处理就是利用粗糙集理论的概念和方法对神经网络的输入数据进行预处理，[6] 具体过程如下：外部数据获取，产生决策表：将鱼病案例数据分为两部分，其中 80 个案例作为训练数据，其余 47 个作为测试数据。对数据再进行如下处理。

（1）空值处理

对于案例数据集中的空值利用均值/众数法进行补偿，即对于连续型属性的空值，用该属性的非空值数据的均值来代替；对于字符型或离散数值型属性空值，空值用该属性数据中的众数代替，即属性值中出现频率最高的值。

用 $a(x)$ 和 $a\hat{}(x)$ 分别代表 x 对象在作空值处理前后对应的 a 属性的值，$O_a=\{x\in U:a(x)\neq *\}$，其中 $*$ 表示空值，$O_a^v=\{x\in O_a:a(x)=v\}$，其中 v 为非连续型值，则有式(17-1)所示关系。

$$a\hat{}(x)=\begin{cases}a(x), & \text{如果 } x\in O_a\\ \dfrac{1}{|O_a|}\sum\limits_{x\in O_a}a(x), & \text{如果 } x\notin O_a \text{ 且对象 } x \text{ 对应的 } a \text{ 属性的值为连续型}\\ \arg\max\limits_{v}|O_a^v|, & \text{如果 } x\notin O_a \text{ 且对象 } x \text{ 对应的 } a \text{ 属性的值为非连续型}\end{cases}\tag{17-1}$$

一般有空值的属性都要作属性的均值或众数的计算，当用户向系统输入信息时，也可能出现某属性值输入为空，这时也需要使用该属性的均值或众数作为其默认值。

（2）属性离散化

对于连续型或整数型变量，可以先用自然算法得到备选分割点，再用布尔推算法合并一些分割点，进行离散化。用布尔推算法不能实现离散化，则可以考虑用等频法进一步作离散化处

理。这里由于只有温度属性理论上属于连续型变量,而实测值都是离散值,因此直接采用离散的实测值作归一化处理即可。

(3) 名词变量的二值化

设置与名词变量值的个数相同的单元进行二值化处理,例如,对于"季节"属性有4个可取值,则设置4个不同的属性S1、S2、S3和S4来代替"季节"属性,如当"季节"属性取值为"春"时,S1的取值为1,S2、S3、S4的取值为0。这样处理后不但可以利用约简算法对新的二值变量(属性)进行化简,而且还可以将数据直接输入神经网络进行训练。名词变量的二值化结果如表17-2所示,对于本身是二值的属性,如各种症状的有无等,无需再作处理。

表17-2 预处理后的鱼病案例决策表

案例编号	条件属性													决策属性(疾病)				
	S1	S2	S3	S4	T1	T2	T3	T4	Temp	A01	A02	…	H14	D28	D31	D35	D36	D40
1	1	0	0	0	0	1	0	0	0.10	1	0	…	0	1	0	0	0	0
2	1	0	0	0	0	1	0	0	0.14	0	1	…	0	0	1	0	0	0
3	1	0	0	0	0	1	0	0	0.07	0	0	…	0	1	0	0	0	0
4	1	0	0	0	0	1	0	0	0.05	0	0	…	0	0	1	0	0	0
5	1	0	0	0	0	1	0	0	0.04	1	0	…	0	0	0	1	0	0
6	1	0	0	0	0	1	0	0	0.11	0	0	…	0	0	1	0	0	0
7	1	0	0	0	0	0	1	0	0.12	1	0	…	0	1	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
125	0	1	0	0	0	0	1	0	0.15	0	0	…	1	0	0	1	0	0
126	0	1	0	0	0	0	1	0	0.18	0	0	…	0	1	0	0	0	0
127	0	1	0	0	0	0	1	0	0.21	0	0	…	0	0	0	0	0	1

(4) 数据转换

这一步是将原始案例数据转换为适合输入神经网络的形式。对于名词变量已经作了二值化处理,可以直接输入网络。对于其他变量,如连续型变量还需作诸如均值-标准差变换,再作为输入神经网络的数据进行训练或测试。经上述处理后的决策表见表17-2。

(5) 属性约简

经二值化处理后,条件属性总数变为93个,利用前面介绍的属性约简算法求得最小约简为{S1,A02,H01,H03,…},该约简的条件属性仅为22个,将它们作为输入神经网络的变量,构造待训练神经网络。

2. 神经网络学习

知识获取的第二阶段是神经网络的训练,这里利用蚁群算法来完成。[2]

(1) 测试数据的准备

对于47个原始测试数据,如果含有空值,则用从训练数据得到的相应属性的默认值填补;利用训练数据的离散化结果对测试数据进行离散化,并二值化名词属性;选择训练数据的属性约简作为测试集的约简;这样就得到神经网络的测试数据。



（2）网络参数设置

选取三层前馈网络模型，与约简后的变量相对应，网络的输入节点为 22 个，1 个隐层，隐层节点数按经验公式 $\sqrt{mn}$ 计算并经实验确定为 9 个。隐层节点激励函数选为对称 sigmoid 函数，即

$$f(x) = -1 + \frac{2}{1+\exp(-x)} \tag{17-2}$$

输出层节点 5 个，激励函数选为 sigmoid 函数，即

$$f(x) = \frac{1}{1+\exp(-x)} \tag{17-3}$$

误差限取为 0.001，批的规模为全部训练数据。蚁群算法参数选为 $\rho=0.7$，$s=50$，$Q=60$。神经网络参数 p_i 为 −1.8 和 1.8 之间的随机数，N 取 60。

（3）训练结果

当选择阈值为 0.25，即满足 $|o_i-d_i|<0.25$（其中 d_i 为标准输出，o_i 为实际输出），训练 2430 次得到对测试集正确率为 94.23%，当选择阈值为 0.30 时，对测试集正确率为 96.44%。

利用训练好的神经网络作为系统的知识库，即把满足训练要求的权值矩阵、神经网络结构参数、网络数据标识信息存入数据库，待推理使用。

本系统实际上经过对 8 个缩简后的案例空间分别作上述处理，最后获取 8 组神经网络知识数据。在推理时，根据输入数据中属性的分布情况优先选择某组知识数据作推理。

17.3.3 推理机制

BP 网络的推理机制采用数据驱动策略，具体步骤如下。

（1）将数据提交给输入层各节点；

（2）由 $x_i=\sum\limits_j \omega_{ij}y_j+\theta_i$，其中 $0<\theta_i\leqslant 1$，ω_{ij} 为通过学习得到的网络权值，计算输入层单元的输出 $y_i=f(x_i)$，并将其作为隐层神经元的输入；

（3）由 $y_i=-1+2\Big/\Big(1+\exp\Big(-\Big(\sum\limits_j \omega_{ij}y_j+\theta_i\Big)\Big)\Big)$ 计算隐层节点的输出，作为输出层神经元的输入；

（4）由 $y_i=1\Big/\Big(1+\exp\Big(-\Big(\sum\limits_j \omega_{ij}y_j+\theta_i\Big)\Big)\Big)$ 计算输出层神经元的输出；

（5）由阈值函数判定输出层神经元的输出，即 $O_i=\begin{cases}\text{yes}, & |y_i-d_i|\geqslant 1-\alpha\\ \text{no}, & \text{其他}\end{cases}$，其中 d_i 为标准输出，α 为设定的阈值。

17.4 系统开发实现举例

（1）根据移动终端的显示特点，三种移动应用方式采用近似的界面形式。以运行于索爱 P800 系列智能手机上的嵌入式鱼病专家系统为例，图 17-6 和图 17-7 给出了几个界面实例。

（2）知识获取系统、普通 Internet 浏览器使用实例，如图 17-8 和图 17-9 所示。



(a) 目检——选择症状图



(b) 目检—— 诊断结果

图 17-6　目检



(a) 远程诊断——问诊信息



(b) 远程诊断 ——专家应答

图 17-7　远程诊断



图 17-8　鱼病诊断知识获取系统——症状提取



图 17-9　鱼病诊断推理结果输出页面

17.5　小结

本章全面分析了目前鱼病专家系统的现状及存在的问题，提出了一种符合用户需求特点的移动式鱼病专家系统软硬件解决方案。对移动式鱼病专家系统的知识获取和推理机制进行了深入研究和设计，实现了基于智能计算技术的知识获取系统和推理机。针对移动式的应用特点，将专家系统分为固定式专家系统和移动式专家系统部分，从而有效解决了用户在不同地

点、不同应用环境下的使用问题。针对移动设备资源有限的特点，充分利用服务器计算资源，把知识获取系统交给服务器完成，再将经简化的神经网络知识库和推理机嵌入移动设备，从而有效保证嵌入式专家系统的诊断能力是一个创新点。

参考文献

1. 郑育红，傅泽田，张小栓．鱼病诊断专家系统设计[J]．中国农业大学学报，2000，5(6)：94～97
2. 胡利平，许永诚，高文等．蚁群神经网络在鱼病专家系统中的应用研究[J]．微计算机信息，2005，11：149～151
3. 胡利平，许永诚，高文等．移动式水产网络化诊断决策系统相关技术研究[J]．邢台职业技术学院学报，2004，21(5)：59～61，69
4. 李松仁，景广军，陈松乔等．基于 multi-agent 的分布式专家系统原理及应用[J]．中南工业大学学报，2001，32(1)：20～24
5. 徐进，郝红艳．Client/Server 体系结构分布式专家系统模型实现[J]．微计算机应用，1997，18(4)：218～221
6. 胡利平，周云，田东等．基于粗分析的案例检索算法在鱼病诊断中的应用研究[J]．计算机工程与应用，2005，18：184～187
7. 于银辉，杨长保，刘志辉．基于多 Agent 的分布式专家系统及其应用[J]．吉林大学学报(地球科学版)，2005，35(2)：266～269

第18章

奶牛疾病诊断专家系统的设计与实现

本章对奶牛疾病诊断专家系统的体系结构、工作流程、各主要功能模块的程序设计进行了说明，重点分析了奶牛疾病诊断系统知识库和推理机的设计。

18.1　系统体系结构设计

基于 Web 的奶牛疾病诊断专家系统的结构分为三层，即数据层、业务层和表现层（见图 18-1）。

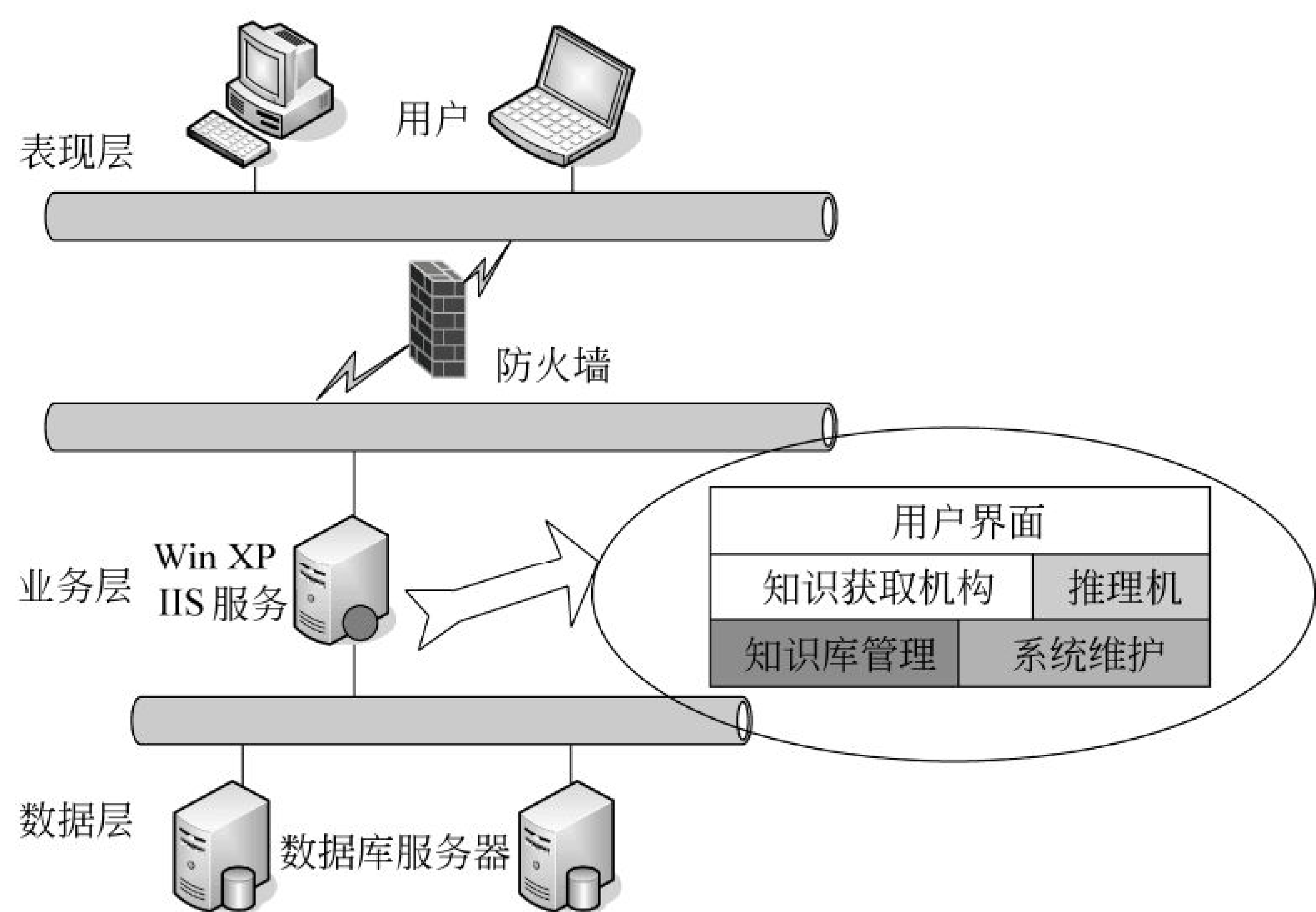


图 18-1　系统体系结构图

数据层：存储在数据库服务器中的主要是知识库和辅助信息库，知识库主要是规则库，辅助信息库中主要是疾病信息、防治信息、药品信息、用户信息以及诊断过程中使用的临时动态表等。

业务层：核心是奶牛疾病诊断专家系统，包括专家系统的一般结构，有用户界面、知识获取机构、推理机、知识库管理和系统维护模块。业务层还包括用户管理、知识浏览、诊断历史查询等功能模块。

表现层：用户通过访问 Internet 来使用本系统，这种基于 Web 的专家系统解决了用户找专家的时间和空间限制的问题，且不需用户手动更新系统和数据。

18.2　系统功能模块设计

本系统的功能模块包括：奶牛疾病诊断模块、知识浏览模块、诊断记录查询模块和系统维护模块，[1] 每个功能模块又由若干个子模块组成，如图 18-2 所示。

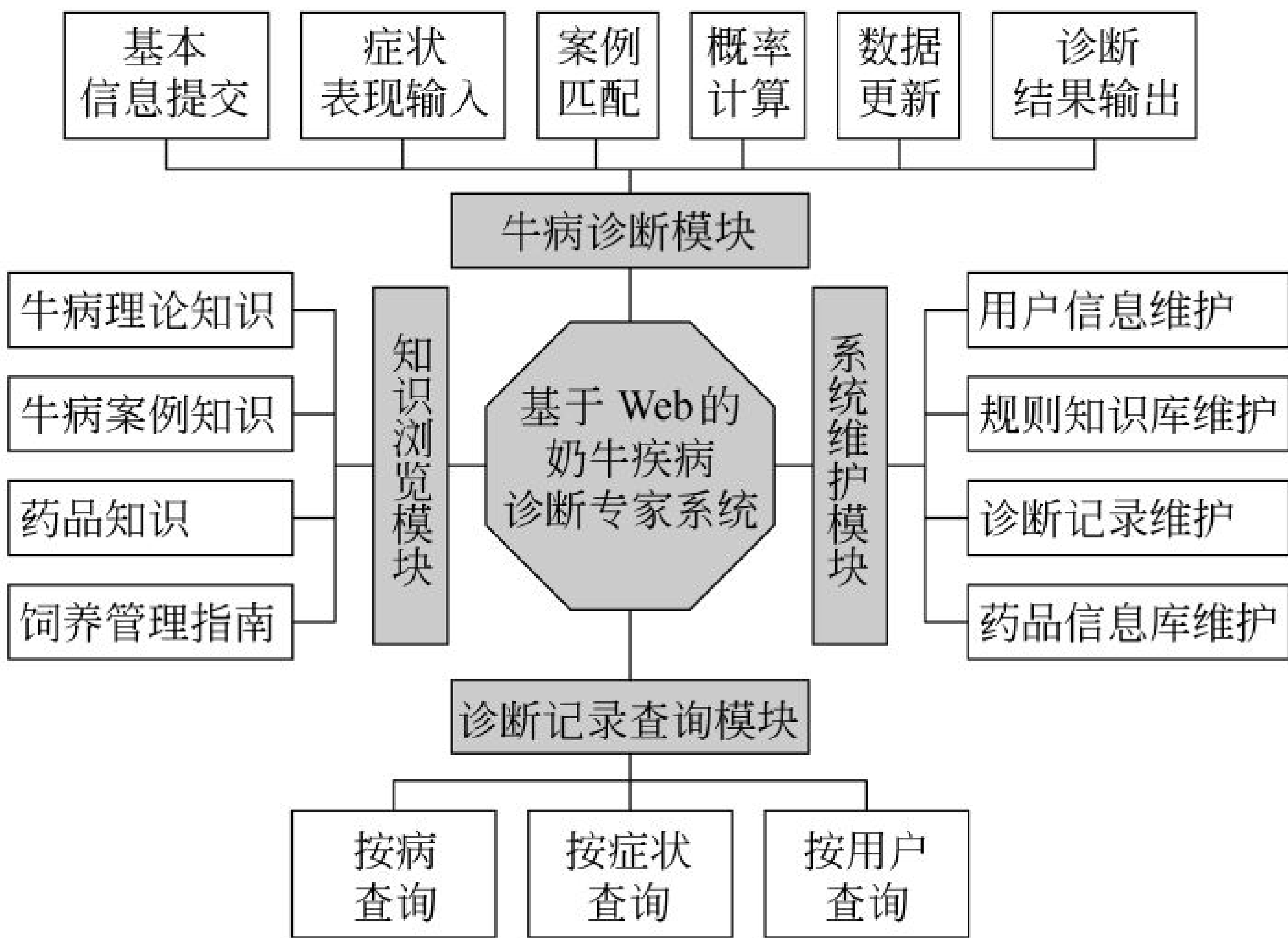


图 18-2　系统功能模块图

(1) 奶牛疾病诊断模块：基本信息提交→症状表现输入→案例匹配→概率计算→诊断结果输出。

(2) 奶牛疾病知识浏览模块：主要介绍奶牛疾病理论知识如病名、症状、病理变化、类症鉴别、防治措施、流行特点等；药品知识包括牛场常用生物药物的介绍、保存及使用方法，牛场常用化学药物的介绍、保存及使用方法等。

(3) 诊断记录查询模块：查询相关信息，主要作为日后诊断的参考，充实知识库。

(4) 系统维护模块：该模块实现对奶牛疾病知识库、诊断记录库、药品信息库和用户信息库的维护，包括数据的添加、修改与删除等操作。

系统功能如图 18-3 所示。

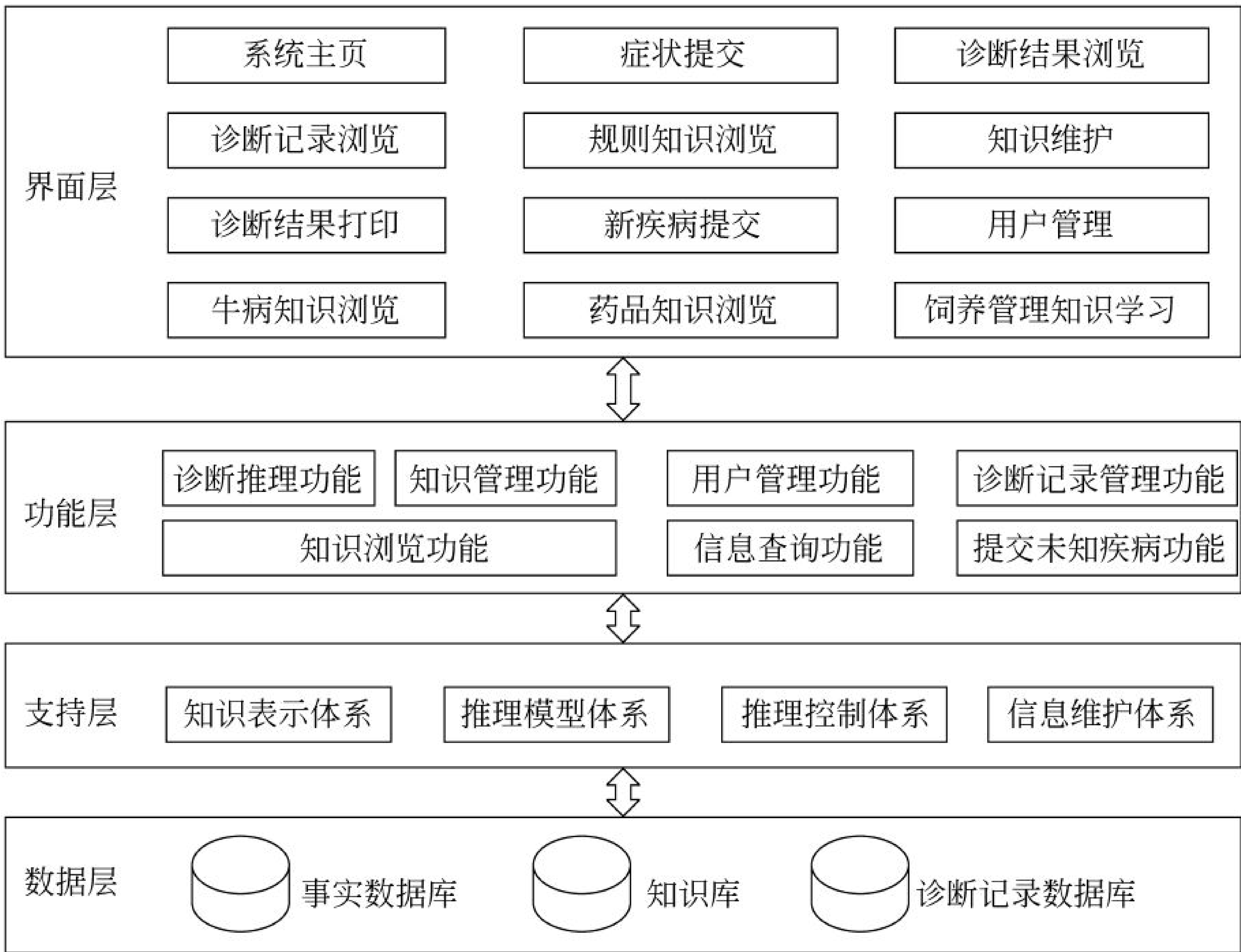


图 18-3　系统功能描述



18.3　系统工作流程设计

系统工作流程如图 18-4 所示，用户首先进入用户登录界面，输入登录信息后，系统根据不同用户的身份赋予不同的权限。普通用户登录后使用诊断系统进行奶牛疾病诊断，或进行奶牛疾病知识浏览，包括奶牛疾病理论知识、奶牛疾病案例知识、药品知识和饲养管理指南等。也可进行诊断记录的查看，用户可以通过奶牛疾病、症状表现和用户三个方面对案例进行查询。[2] 普通用户还可以修改自己的注册信息。

图 18-4　系统工作流程

专家用户登录后可以进行知识库的管理,包括对规则、知识的添加、修改、删除。

管理员登录后可以对系统进行维护,包括对奶牛疾病诊断记录库、知识库、药品信息库和专家信息库的维护,包括数据的添加、修改与删除等操作。

18.4 系统知识库设计

知识实现的主要任务是将形式化的奶牛疾病知识转换为由编程语言表示的可供计算机执行的语句和程序,从而初步完成可执行的奶牛疾病诊断的原型系统。

根据奶牛疾病知识概念化模型和形式化定义的阐述,并根据"方便推理和使用"的原则,将奶牛疾病诊断知识按内涵分为概念、事实及规则等。[3]概念和事实定量或定性地描述奶牛疾病诊断信息,包括症状、疾病、病因、防治的集合信息;规则是反映事实及概念间(症状与疾病、病因、防治集合之间)内在的必然联系,描述为前提条件与结论的因果关系,包括环境条件-疾病库、症状-疾病库、疾病-病因症状库、病因症状-病因库、病因库-防治库,规则库根据数据关系联系在一起,共同完成对奶牛疾病知识的表示。如图 18-5 所示。

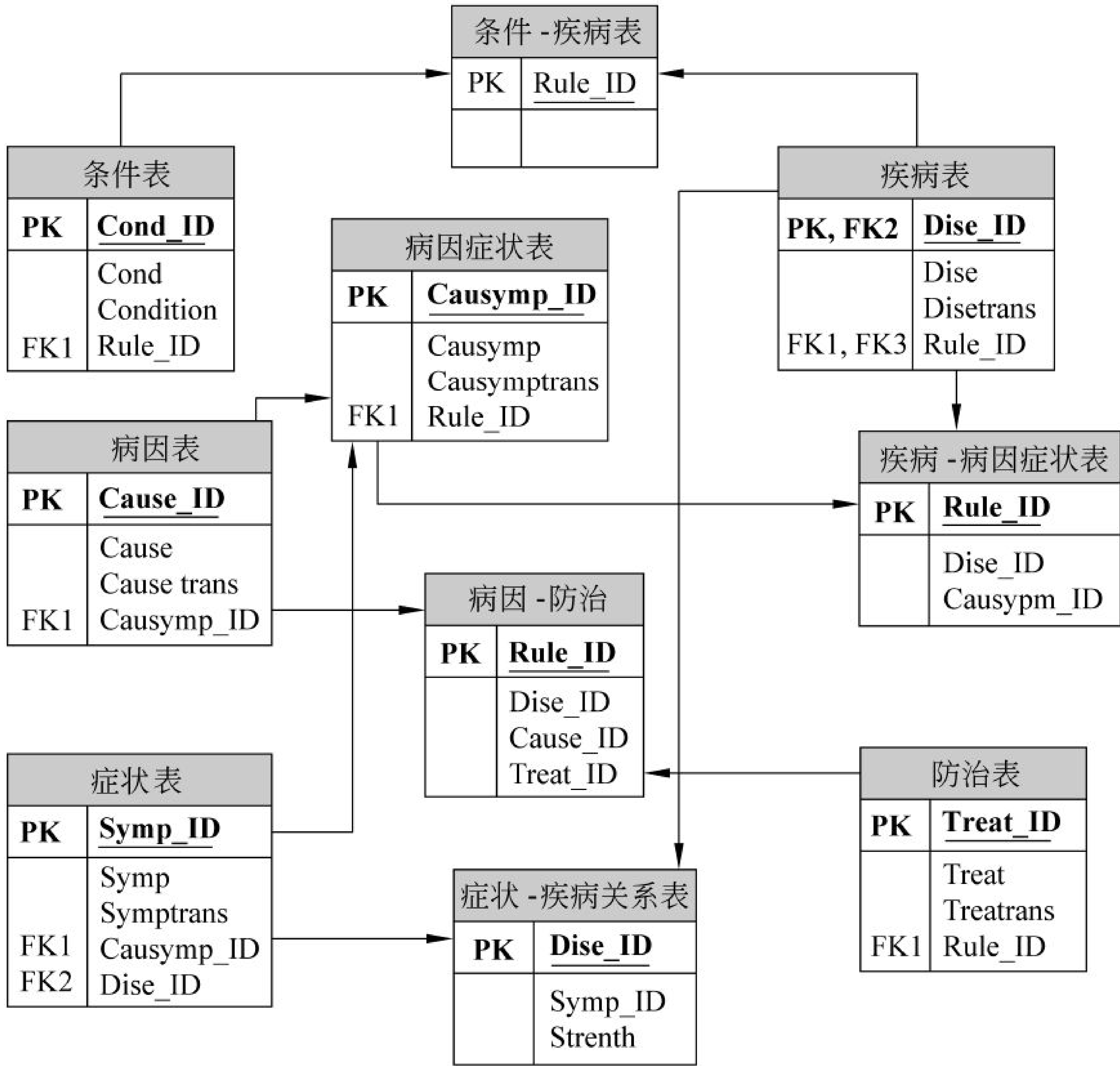


图 18-5 系统知识库设计

18.4.1 字典数据库

奶牛疾病症状、疾病、病因和防治等概念和事实存放在字典数据库中,分别表示为:

(1) 症状表 SYMP_DB(Symp_ID,Symp,Symptrans)

(2) 疾病表 DISE_DB(Dise_ID,Dise,Disetrans)



(3) 病因表 CAUSE_DB(Cause_ID,Cause,Causetrans)

(4) 防治表 TREAT_DB(Treat_ID,Treat,Treatrans)

(5) 病因症状表 CAUSYMP_DB(Causymp_ID,Causymp,Causymptrans)

其中症状表根据皮毛(ZA)、头部(ZB)、运动系统(ZC)、神经系统(ZD)和泌殖系统(ZE)部位编号,如:皮毛共有 28 种特征症状,依次编号为 ZA_1、ZA_2,…,ZA_{28};疾病表根据细菌病(XJB)、病毒病(BDB)、真菌病(ZJB)、寄生虫病(JSB)以及由环境产生的疾病(HJB)编号,如:细菌病细分为 10 种疾病,依次编号为 XJB_1,XJB_2,…,XJB_{10},病因表和防治表同理。

18.4.2 规则库

规则库分为环境条件-疾病库、症状-疾病库、疾病-病因症状库、病因症状-病因库、病因-防治库 5 个规则库,每个规则库均有规则前件数据库(RULE_PRE_DB)和规则后件数据库(RULE_ACT_DB)。[1]

规则前件数据库的结构为:RULE_PRE_DB(Rule_id,Fact_ID,Active),该库用来存放各条规则对应的前提条件。其中:Rule_id 为规则编号,Fact_ID 为 Rule_Name 规则的一个条件,代表上述字典库中的概念或事实,Active 为 Fact_ID 激活的标志,以避免规则各前提条件的重复匹配。一条规则的 n 个与条件在该数据库中就有 n 条对应该规则的记录。

规则后件数据库的结构为:RULE_ACT_DB(Rule_id,Fact_ID,Num,CF,Active)该库用来存放各条规则的结果,其中:Rule_Name 为规则名,Fact_ID 为 Rule_Name 规则的结果,Num 表示其前提条件的个数,CF 为推出 Fact_ID 的可信度,Active 表明本规则是否已经激活,以避免规则的重复激活,从而提高推理效率。通过 Fact_ID 使字典库与 RULE_PRE_DB 中的前提条件关联起来,同时又通过 Rule_id 使 RULE_PRE_DB 与 RULE_ACT_DB 建立起关联。

18.5 奶牛疾病诊断对象的推理机设计

推理机设计主要指诊断推理流程的设计。推理机的诊断算法如下:

步骤 1:检查是否有相应案例,若无则进入步骤 3;

步骤 2:案例的相似度计算,若达到案例激活阈值,则进入步骤 7;

步骤 3:检查是否有相应规则,设为 n 条,若 $n=0$ 则进入步骤 4,否则进入步骤 5;

步骤 4:搜索证据,若新证据不为 0,则进入步骤 2;

步骤 5:计算第 i 条规则的激活概率,得到返回值,判断是否满足激活阈值,若否,则 $i=i+1$,重新进入步骤 5;

步骤 6:判断是否 $i=n$,若否,则 $i=i+1$;

步骤 7:得到返回值列表,更新问题对象。

诊断推理的具体流程如图 18-6 所示。

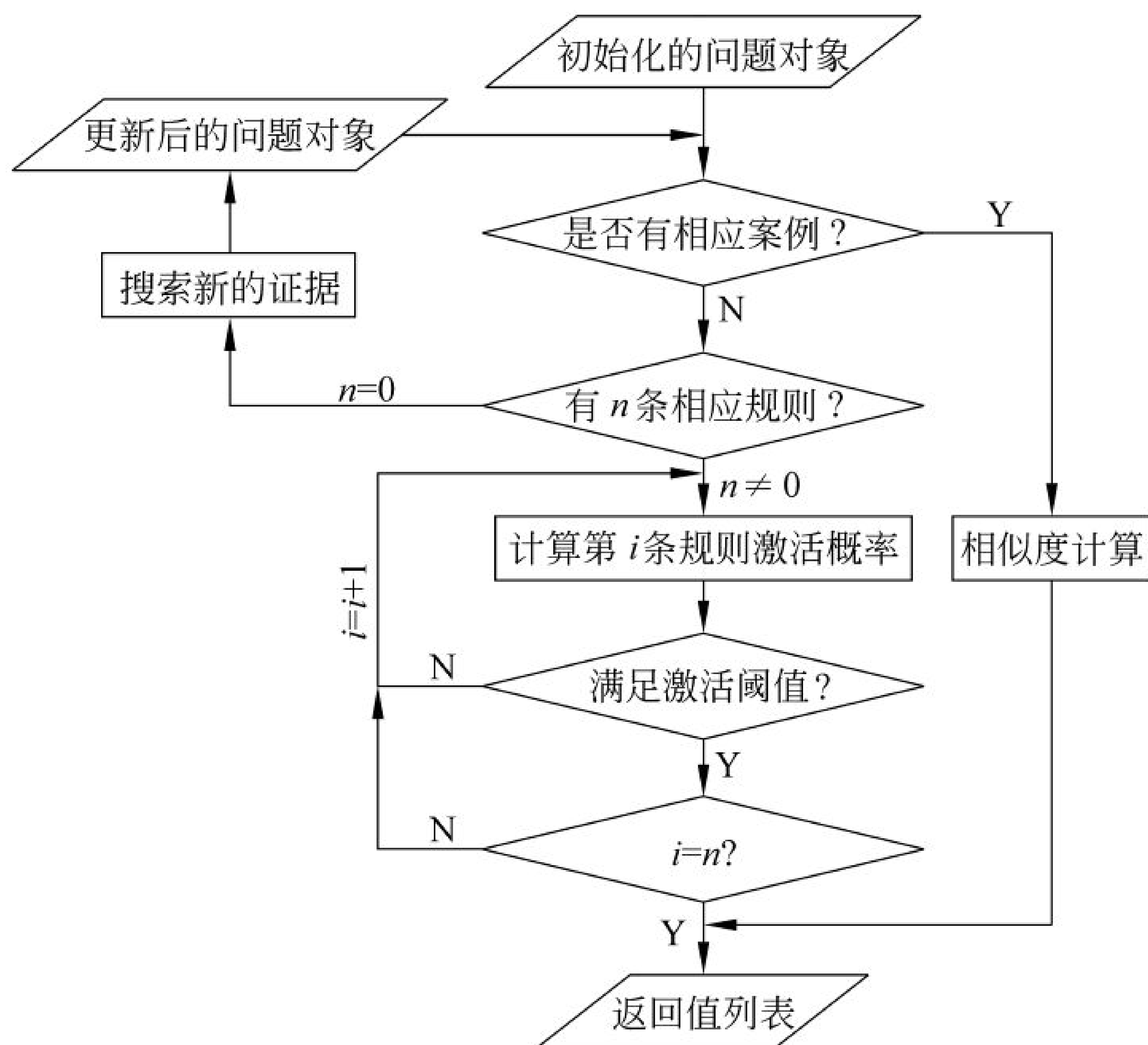


图 18-6　推理机的诊断流程图

18.6　系统界面

系统诊断的过程是用户登录后进入诊断界面提交观察到的症状，待选症状按发病部位或生理系统分类，以树状图的形式给出，方便用户快速选择，选完症状后由用户指定每一症状的观测可信程度，然后点击提交进入系统后台计算阶段，系统计算完成后列出可能结果，以结果为真的可能性由大到小排序，用户可以查看第一结果的详细信息，也可选择进一步诊断或者按系统提示进行特定检查，以最终确定诊断结果。见图 18-7～图 18-9。



图 18-7　选择症状



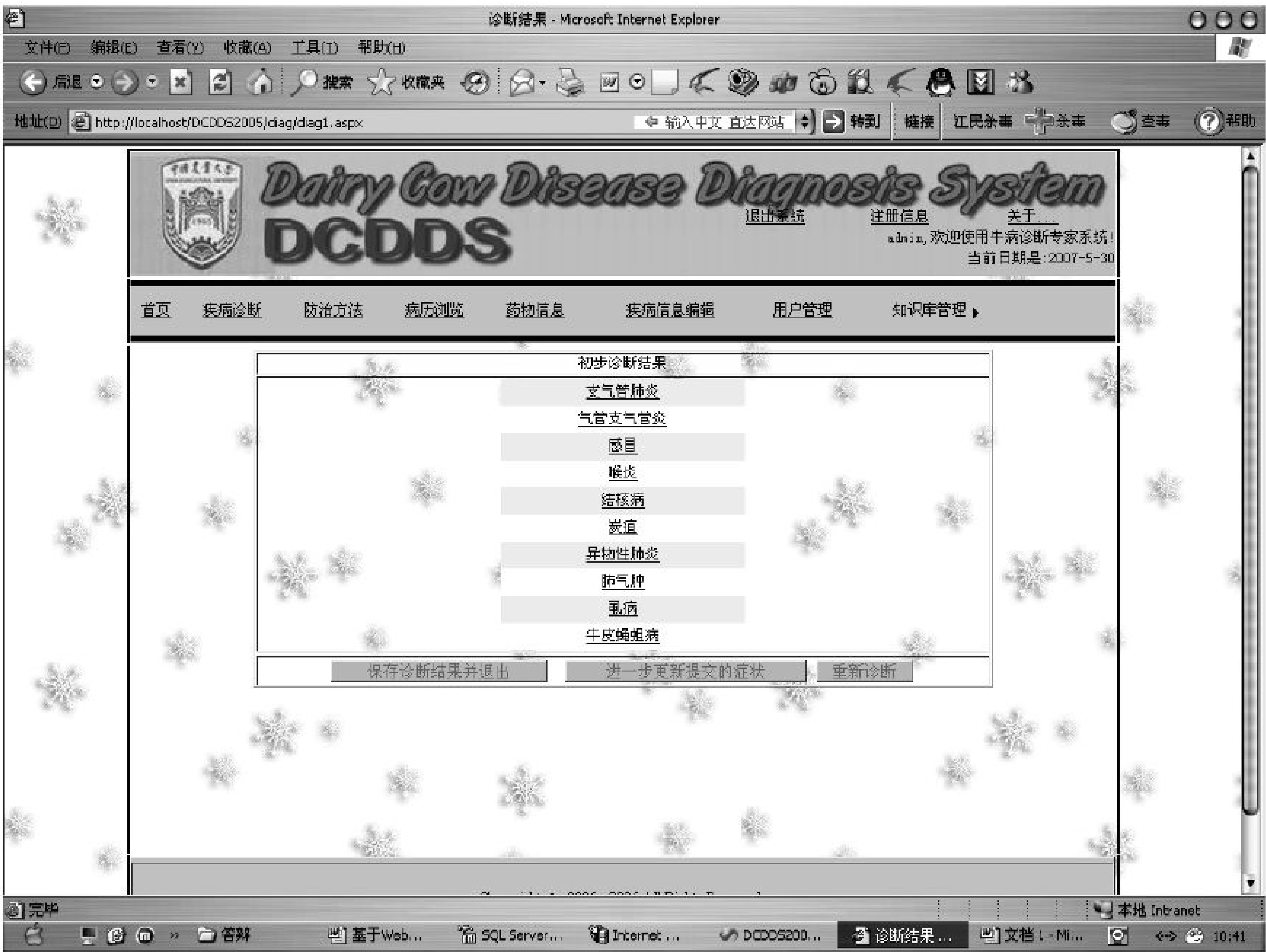


图 18-8 可能结果列表

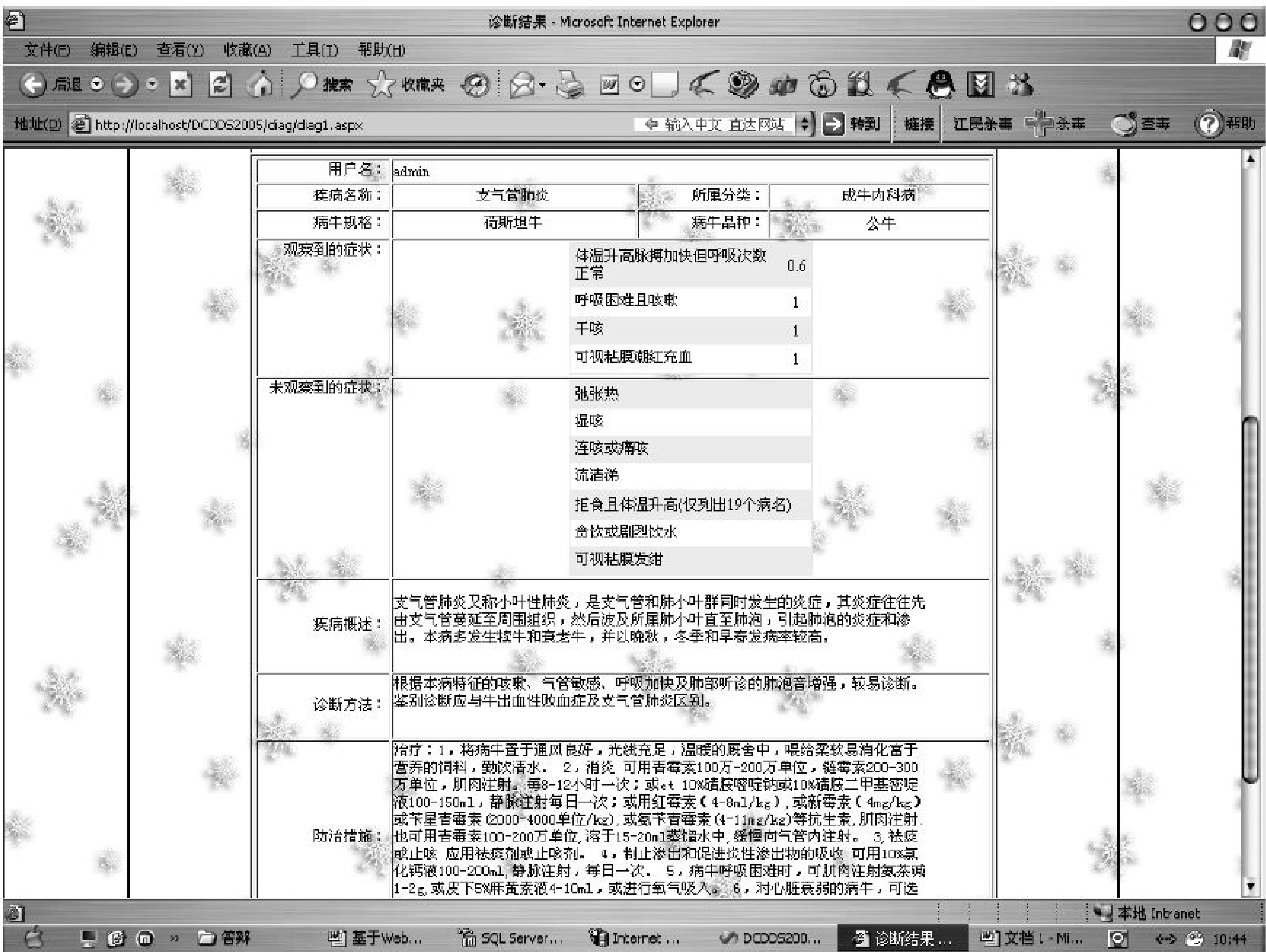


图 18-9 诊断结果验证

18.7 小结

本章采用 ASP. NET 技术开发了基于 Web 的奶牛疾病诊断专家系统，示范结果表明，普通奶牛养殖场可以具备网络条件，很容易登录并使用本系统。由于本系统的系统维护由管理员在服务器端进行，知识库的维护和更新由领域专家负责，减轻了用户的维护和升级负担，受到用户欢迎。

参考文献

1. 戎立斌. 基于 Web 的奶牛疾病诊断专家系统研究[D]. 中国农业大学，2007
2. 李道亮，戎立斌，胡立娟等. 基于 Web 的牛病诊断专家系统的设计与实现[C]. 中国畜牧兽医学会信息技术分会成立大会，2008
3. Rong L , Li D. DCDDS: A Dairy Cow Disease Diagnosis System for Dairy Farm in China[C]. IEEE. 2006

第19章

基于Web的果树病害诊断系统

本章以果树病害诊断问题的特点为基础，按照知识工程的开发方法，并参照软件工程的方法进行研制开发，将果树病害诊断专家系统的开发过程分为：系统分析、系统设计、知识获取、系统实现和系统测试。

(1) 系统分析阶段：主要通过调研、本领域科研积累、文献分析等手段，确定果树病害诊断的总体思路，即先分辨病虫害再分别进行侵染性病害、非侵染性病害的分辨，并将营养失调病害和叶片养分诊断相结合的流程；再根据Web的发展现状，选定PAID作为系统开发平台。

(2) 系统设计阶段：是基于果树病虫害诊断原理，应用系统工程的思想和方法，把复杂的果树病害诊断专家系统模型分解成科学、实用、简单的模块，并确定各模块的具体流程。

(3) 知识获取阶段：主要基于已有的科研成果、文献分析等，按照知识模型(系统设计拟定的流程)，获取果树病害诊断问题求解时所需要的各种知识，并按照开发平台要求转化成相应的知识形式，该阶段的工作为系统开发提供了知识规则、图片、备注文档等基础资料。

(4) 系统实现阶段：主要目标是基于系统设计、知识规则、图片、备注文档等，采用PAID+SQL Sever，快速实现基于Web的果树病害诊断专家系统，并给出详尽的系统界面说明。

(5) 系统测试阶段：主要采用专家评审、模拟测试、实地测试等方法，对系统进行全面检验，修正错误，根据用户的需求动态调整系统，为进一步推广应用打下基础。

19.1 基于Web的果树病害诊断系统需求分析

病虫害诊断与防治是果树无公害生产的关键环节之一。由于误诊或防治措施不当(例如施用农药不当造成的农药残留)，造成果树减产或果实品质下降的现象时有发生，其主要原因在于许多果农缺少科学有效的病虫害诊断与防治技术，同时由于植保专家和基层技术人员相对缺乏，果农得不到及时有效的科学指导。[1]结合对山东烟台、临沂、肥城等地果树产业的定期走访和深入调研，以及对于果树病害诊断领域文献的大量阅读和分析，确定了系统构建总体思路，即先分辨病虫害后诊断。具体叙述如下。

病害诊断的第一步，是首先要能够正确划分病害的类别。植物病害的种类可以有多种分类方法，例如按照作物类型、寄主受害部位、病原生物类型、传播方式或介体等，但最客观也最实用的还是按照病因类型来区分的方法，它的优点是既可知道发病的原因，又可知道病害发生

的特点和防治的对策等。

根据这一原则，植物病害分为两大类，第一类是由病原生物因素侵染造成的病害，称为侵染性病害，因为病原生物能够在植株间传染，因而又称传染性病害，侵染性病害又根据病原生物的类别分为真菌病害、细菌病害、病毒病害、线虫病害和寄生性植物病害等；另一类是无病原生物参与，只是由于植物自身的原因或由于外界环境条件的恶化所引起的病害，这类病害在植株间不会传染，因此称为非侵染性病害或非传染性病害，引起非侵染性病害的病原因子有很多，主要可归为营养失调（缺素症或营养元素过量）、水分失调、温度不适、有害物质（药害、肥害、环境污染等）、遗传性病害、气象灾害（如风、雨、雷电、冰雹等）和农事操作不当所致病害等。[2] 然后根据不同的病害种类，制定不同的诊断规则。

19.2 基于 Web 的果树病害诊断系统设计

19.2.1 系统结构设计

果树病害诊断专家系统是基于 Internet 运行环境开发的，采用了当前流行的"Web 浏览器/Web 服务器/数据库系统"网络体系结构（也称 B/S/S 结构），系统按功能划分为客户层、应用层和数据层三大块。从 ASP. NET 系统结构出发，设计果树病害诊断系统结构如图 19-1 所示。

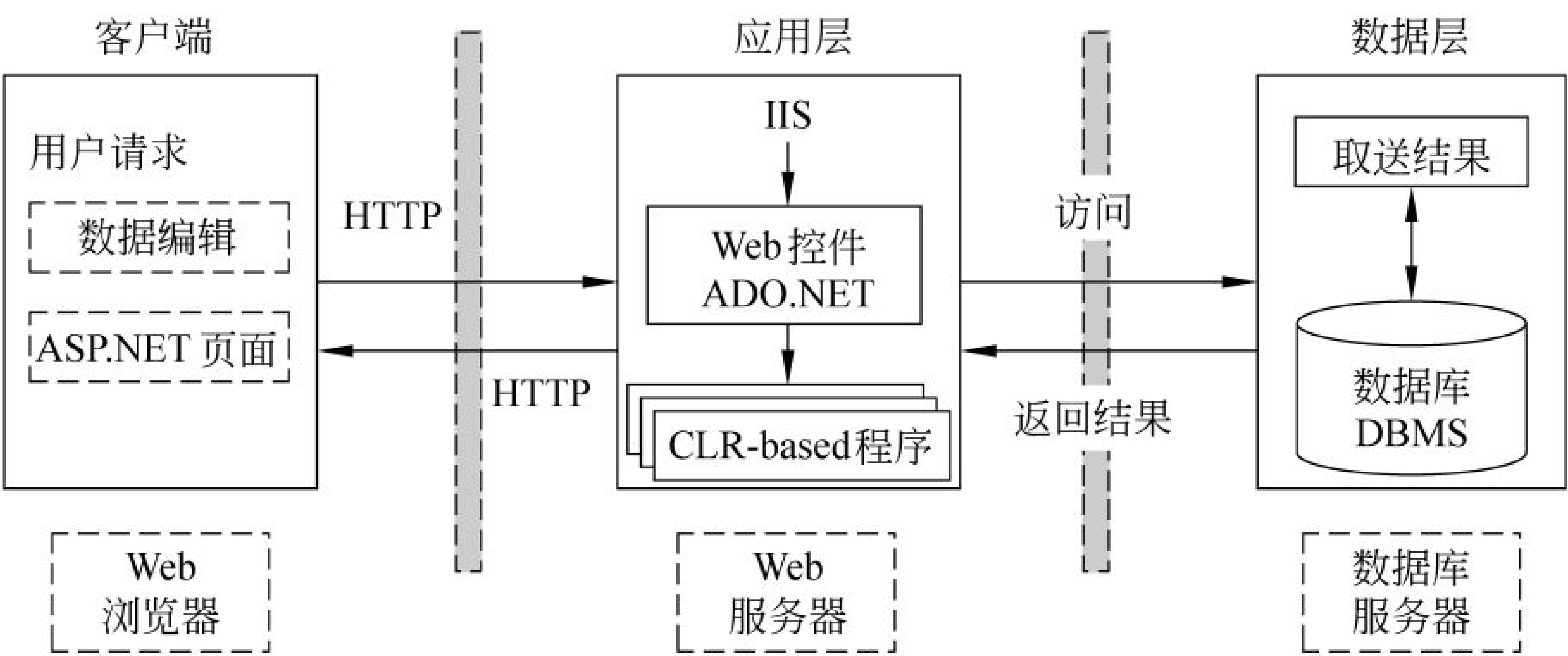


图 19-1 基于 Web 的果树病害诊断系统的三层体系结构图

19.2.2 功能模块设计

基于需求分析和领域知识的特点，设计系统功能模块如图 19-2 所示。

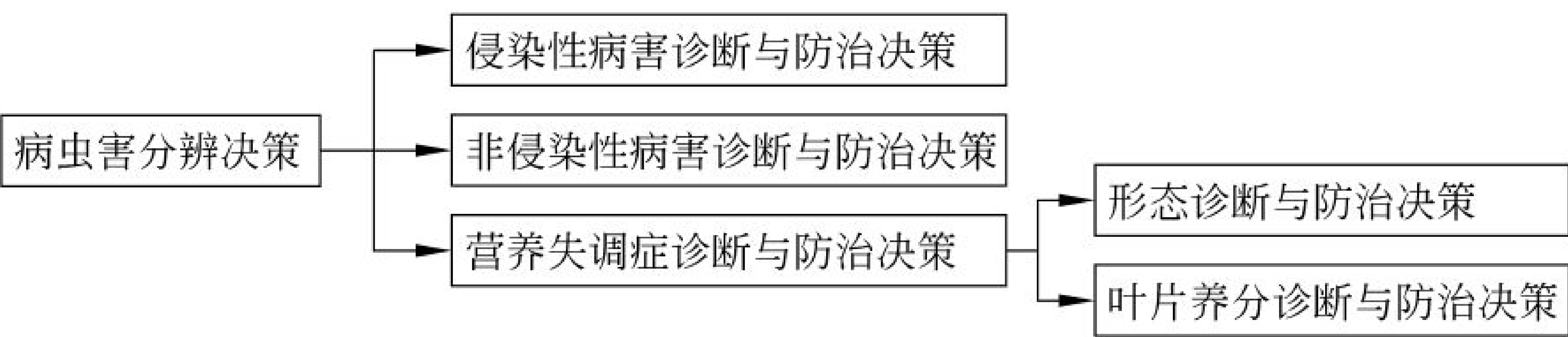


图 19-2 果树病害诊断系统功能模块图



19.2.3　工作流程设计

1. 病虫害分辨决策

根据侵染性病害、非侵染性病害和虫害发生规律，制定出病虫害分辨检索表。

2. 侵染性病害诊断与防治决策

根据侵染性病害的发病部位、症状特点(包括病症和病状)、发病规律等诊断出病害种类，并提供给用户不同部位、不同时期的彩色病害照片以查看确认。综合防治方法包括植物检疫、选育和利用抗病品种、农业防治、化学防治、生物防治、物理防治等，流程见图 19-3。

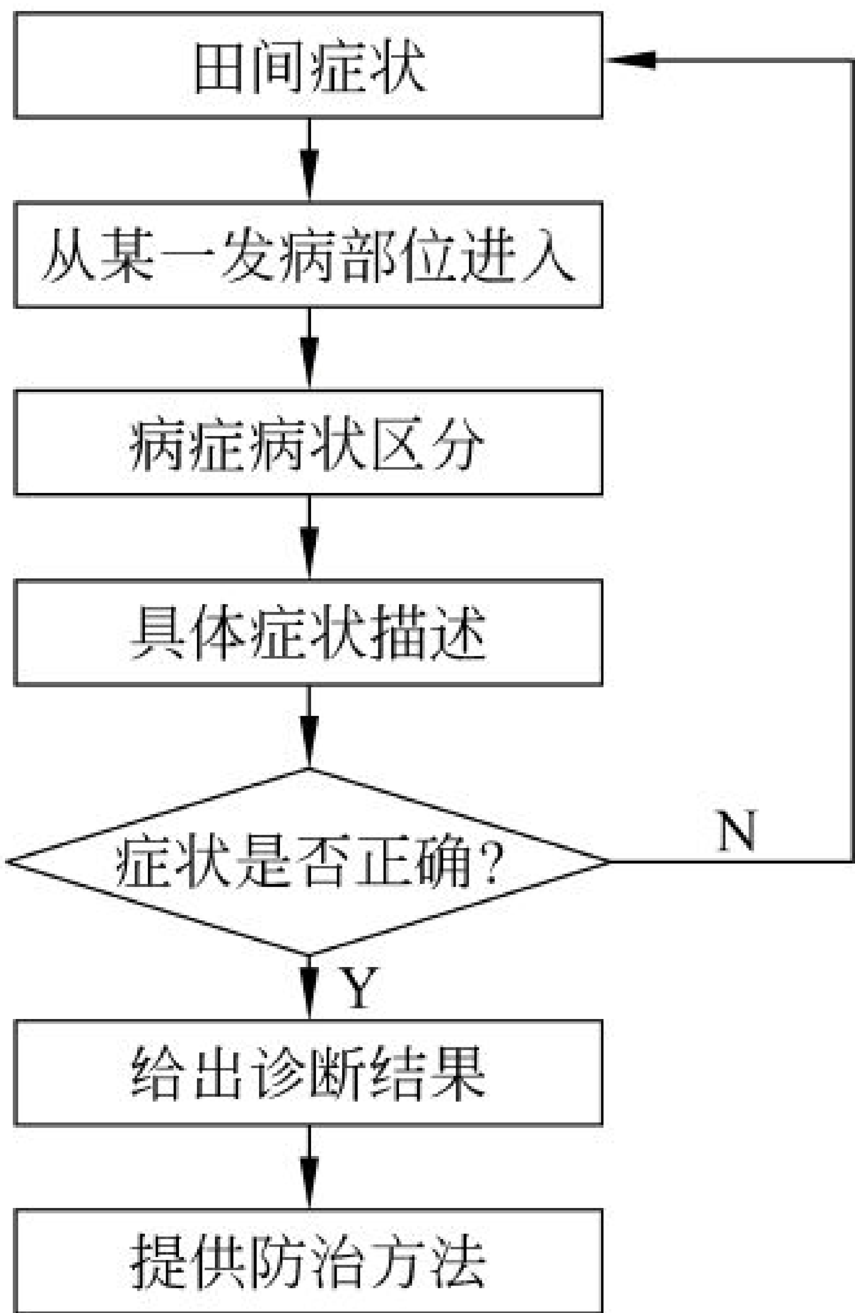


图 19-3　侵染性病害诊断与防治模型框架图

(1) 田间症状

田间症状是指用户在田间实地观察所得到的病害发生情况。由于用户文化水平及语言上的差异，不同用户对同样症状的描述可能不同。为此，本模型采用了一套标准的描述语言，用户把自己观察的情况与系统描述语言相对应录入即可。

(2) 发病部位

采用的病害分类方法是按照植株受害的器官类别划分的。植株各种器官的结构和功能有较大的差别，以致危害的病害种类、发生规律和防治方法都有不同。所以，例如对肥城桃而言，将侵染性病害划分为根部病害、枝干病害、叶部病害、果实病害、花部病害等，依据树种不同可灵活掌握。用户在录入事实时，可以根据自己的观察角度，选择某一发病部位进入。

(3) 病症病状区分

症状是植物生病后的不正常表现，其中寄主植物本身的不正常表现称为病状，病原物在病部的特征性表现称为病征。具体说来，病征有粉状物、霉状物、点状物、颗粒状物、脓状物等，病状有变色、斑点、腐烂、萎蔫、畸形等，以上作为模型基本组件，可针对不同植物和病害种类选择使用。

(4) 具体症状描述

对各种病害的发病部位和发病时期症状表现等进行详细描述。同一种病害可能在不同的部位发病，如细菌性穿孔病可在叶片、枝条、果实上发病，病程一般分为初期、中期、后期或类似的若干时期，各时期症状表现也不尽一致。通常，具体症状描述分为 3～5 步，通过具体症状的描述，可使用户对照田间实际症状，逐步诊断。

(5) 症状确认

用户在录入完数据后，保存决策，系统会自动给出结论和备注文档。为了验证结论的准确性，用户可先打开备注文档进行“确认”：文档前一部分为病害的症状特点、病原和发生规律，并有相应图片供参考，用户可将自己观察到的情况与之相对照，确认无误后，可查看后一部分的防治方法；如认为有误，可重新回到数据录入部分，换一条途径继续诊断。

(6) 诊断结果

在正确诊断之后,系统会提供病害名称及结论的可信度。

(7) 防治方法

按照无公害生产的要求,提供对病害的综合防治方案。

3. 非侵染性病害诊断与防治决策

非侵染性病害只有病状而没有病征,多由水分供应失调、温度过高或过低、空气中有毒气体的存在、农药使用不当、养分不足或不均衡等原因造成,所以要根据以上情况和发病部位、病状特点等进行诊断。如日烧病是由于阳光过度直射等造成的,其诊断流程类似于非侵染性病害。

4. 营养失调症诊断与防治决策

本系统将营养失调症从非侵染性病害中单列出来,提供两种诊断决策模块:一是以形态诊断为主,依据"果树矿物质元素失调症诊断检索表"进行诊断;二是以叶片矿质元素含量为指标的组织分析诊断法,基于果树叶片各必需矿质元素的适宜范围和营养失调的临界浓度诊断。根外追肥诊断法、土壤营养诊断法等归入综合信息查询系统。

(1) 形态诊断与防治决策

以形态诊断为基础,从树体缺素病状表现分析入手,诊断缺素的种类,提出防治方案,包括该元素的生理作用、失调症状、发生规律和防治措施(包括应急措施和根本措施)等在内的浏览文档,用于诊断核实(见图 19-4)。

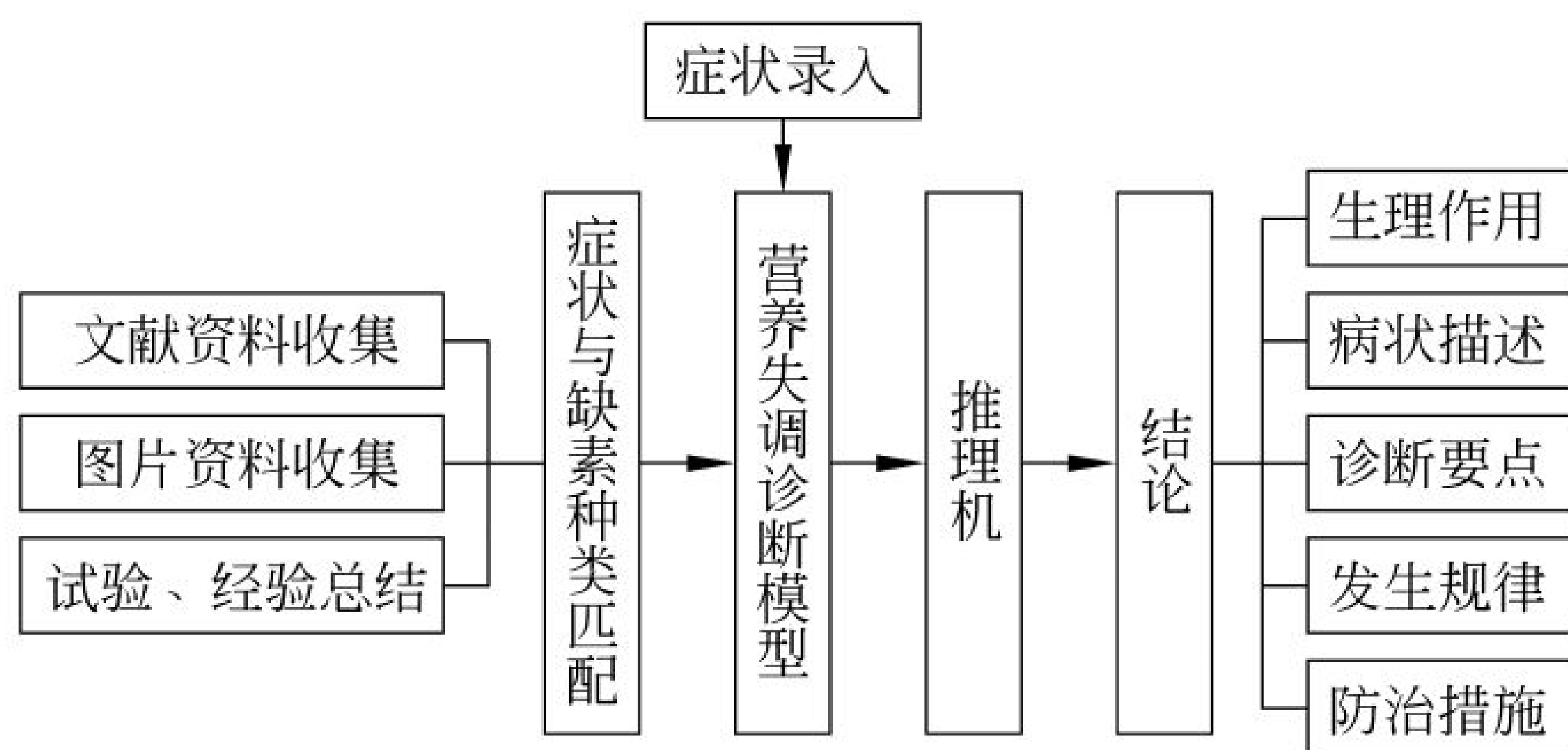


图 19-4 果树营养失调症形态诊断流程图

(2) 叶片养分诊断与防治决策

作为营养失调症形态诊断的补充和确切判断树体营养水平的另一途径,叶片养分诊断是按照正确的取样方法,通过测定叶片中的养分含量,配合专家系统诊断,与正常叶片的养分含量水平相比较,判断元素含量丰缺。如得出营养失调症,则给出其防治方案。元素选择顾及大(常)量元素与微量元素两层面,涵盖对植株生理影响较大、用仪器较易测定的 10 种大(微)量元素。大量元素(干重%):N、P、K、Ca、Mg;微量元素(mg/kg):Fe、Mn、Cu、Zn、B。



19.3 知识获取

19.3.1 知识来源

我们按照果树生长、环境条件和措施效应为有机整体系统的思想，基础理论研究与应用技术紧密衔接的研究方法，科学研究与试验、示范、推广相互结合的技术路线，先后开展了与肥城桃无公害优质丰产栽培的植物生理学、气象学、土壤肥料学、病虫草害无公害防治、贮藏加工等方面相关的 34 项科研课题的系统研究；组织了 300 多项次不同地区、不同条件下多年、多点的系列优质丰产栽培措施和综合高产技术联合试验；连续多年系统观察、解剖分析了肥城桃不同树龄、不同品种类型、不同物候期各器官的生长发育过程和各器官之间的相互关系，摸清了肥城桃优质原因、生长发育全过程的基本规律及其在不同环境条件下的变化特点和栽培技术措施所产生的综合效应；广泛系统的不断进行生产考察和实地调查，掌握了大量不同立地条件、不同气候环境和不同时期的生长发育特点、树势树相、产量品质、水分和养分变化动态等技术资料，积累了大量的第一手科学试验数据。全面深入地掌握了肥城桃生长发育、环境条件和栽培措施相互之间错综复杂的矛盾统一关系，创建了以园地选择与规划、无公害平衡施肥、节水灌溉、整形修剪调控、病虫草害诊断与无公害防治模式为核心内容的“肥城桃无公害优质丰产栽培标准化生产技术体系”，并在全市各肥桃适宜区推广应用，从理论到实践实现了预见性、主动性和应变性，并实现了栽培学科研从单项到综合、从静态到动态、从定性到定量的转变，形成了巨大的生产力，为广泛的知识获取，知识库的建立和开发，形成最终的智能决策系统提供了可靠的保证。

19.3.2 知识模型

基于上述科研成果，借鉴 CommonKADS 方法，开发出果树病害诊断与防治知识模型框架，这样便于知识规则的提取和编辑。依据模型框架和开发平台特点，选择具体知识，将知识规则用产生式表示。例如，针对桃细菌性穿孔病，霉斑穿孔病，褐斑穿孔病三种病害，按照模板要求“填空”，生成如图 19-6 所示的知识规则。将症状描述拆解开来，提供多种途径进行诊断，例如本例就可从“病斑”和“穿孔”两条途径进入。

19.3.3 知识规则形式化

知识规则编辑完毕后，根据专家系统开发平台的要求，将知识规则形式化。以图 19-6 中的一条规则为例：

如果：部位＝叶部(0.2)；

并且：病症病状区分＝病斑(0.2)；

并且：症状 1＝褐色(0.2)；

并且：症状 2＝穿孔边缘整齐(0.2)；

并且：症状 3＝无坏死残余组织；病斑褐色，潮湿时叶背长污白色霉状物(0.2)。

则：霉斑穿孔病。

备注：霉斑穿孔病病原、发病规律及防治方法。

规则可信度：0.8

Ⅰ　叶部
　Ⅱ　穿孔
　　Ⅲ　穿孔边缘整齐
　　　Ⅳ　无坏死残余组织；病斑褐色，潮湿时叶背长污白色霉状物 ……………… 霉斑穿孔病
　　　Ⅳ　有明显的坏死残余组织；病斑中部淡褐色，边缘紫色或红褐色 ………… 褐斑穿孔病
　　Ⅲ　穿孔边缘不整齐
　　　Ⅳ　多在叶脉两侧或叶片边缘发生，病斑紫褐色至黑褐色，外围有黄晕…… 细菌性穿孔病
　Ⅱ　病斑
　　Ⅲ　褐色
　　　Ⅳ　穿孔边缘整齐
　　　　Ⅴ　无坏死残余组织；病斑褐色，潮湿时叶背长污白色霉状物 ………… 霉斑穿孔病
　　　　Ⅴ　有明显的坏死残余组织；病斑中部淡褐色，边缘紫色或红褐色 …… 褐斑穿孔病
　　　Ⅳ　穿孔边缘不整齐
　　　　Ⅴ　多在叶脉两侧或叶片边缘发生，病斑紫褐色至黑褐色，外围有黄晕　… 细菌性穿孔病

图 19-5　侵染性病害诊断与防治知识规则模式

条件阈值：0.5

其中，条件项括号中的 0.2 为逻辑子式权重，规则可信度为该条规则对事物内在规律的把握程度，条件阈值为触发该条规则并产生结果的最小条件真度。

19.4　系统实现

19.4.1　系统开发平台

系统采用 PAID5.0＋SQL Server 2000 的开发环境。PAID5.0 平台采用“客户层/业务逻辑层/数据库层”分布式结构体系，只要遵循本平台技术规范开发的农业专家系统，无需任何修改均能在 Internet 网络环境下运行（见图 19-6）；平台开发出的专家系统具有对输入的数据进

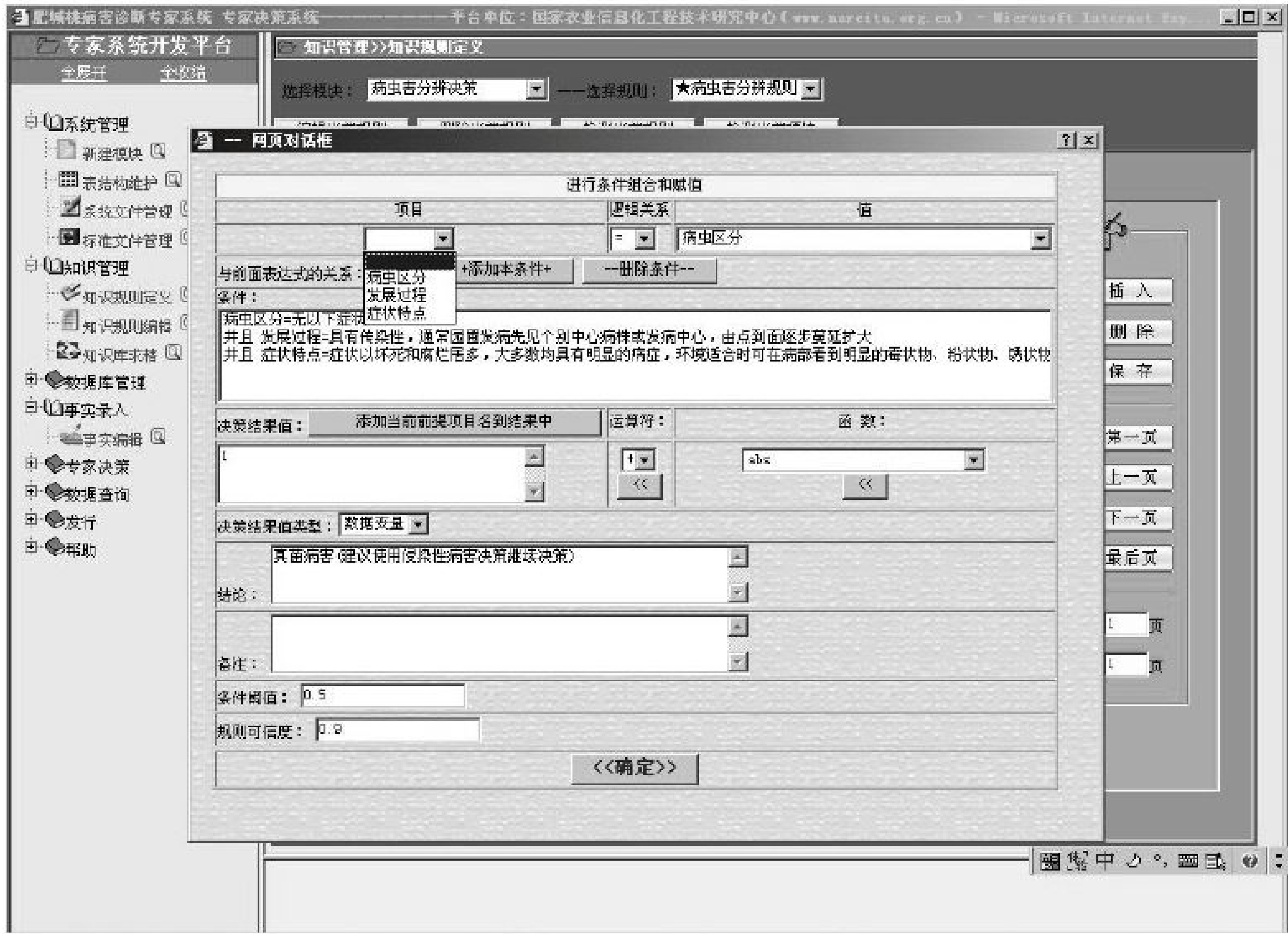


图 19-6　平台知识规则编辑界面



行类型、格式(语法)、范围、数据约束等方面检查的功能,必要时可对输入数据进行形象化解释,确保用户输入数据的正确性。[3,4] 课题组利用该平台开发出了多种果树病害诊断系统,比如肥城桃病害诊断智能决策系统(见图 19-7)。



图 19-7　"肥城桃病害诊断智能决策系统"界面

PAID 系列平台的应用和推广大大缩短了基于 Web 的农业专家系统的开发时间,提高了效率,其流程如图 19-8 所示。[5~7]

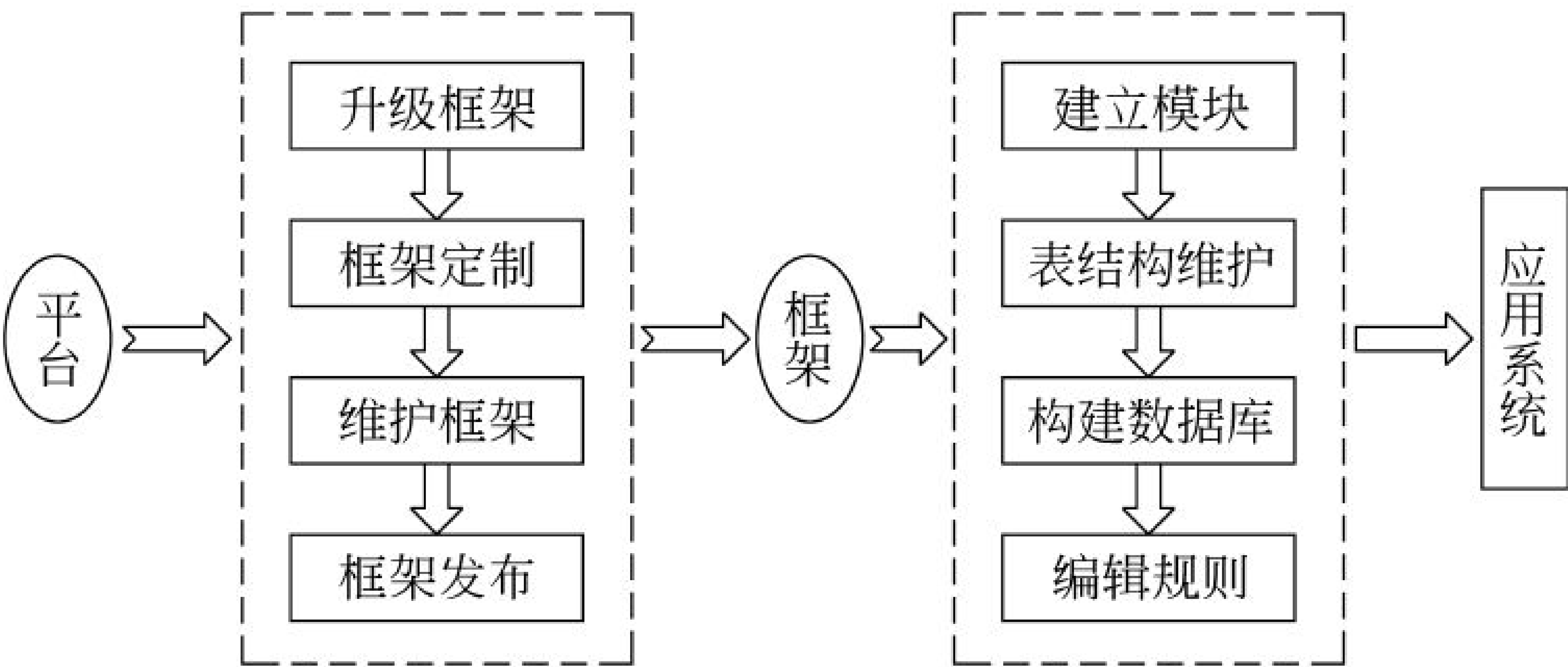


图 19-8　用 PAID5.0 构造农业专家系统的流程图

19.4.2　系统运行界面

1. 病虫害分辨决策

登录系统之后进入系统主页,首先选择左侧"智能决策"中的"在线诊断"链接(见图 19-9),进入"在线诊断"界面,选择"病虫害分辨规则",单击下一步按钮。按步骤依次分别录入病虫区分、发展过程和症状特点(见图 19-9～图 19-11)。录入最后一项之后,单击"下一步"按钮,即自动进行推理并显示结果(图 19-12,图 19-13),使用起来非常快捷、方便。

2. 侵染性病害诊断决策

诊断步骤:首先要选择部位(图 19-14),再录入症状(图 19-15),对最后结果中提供的图片进行确认(图 19-16)。



图 19-9　病虫害分辨事实表录入界面



图 19-10　病虫害分辨事实表“病虫区分”项

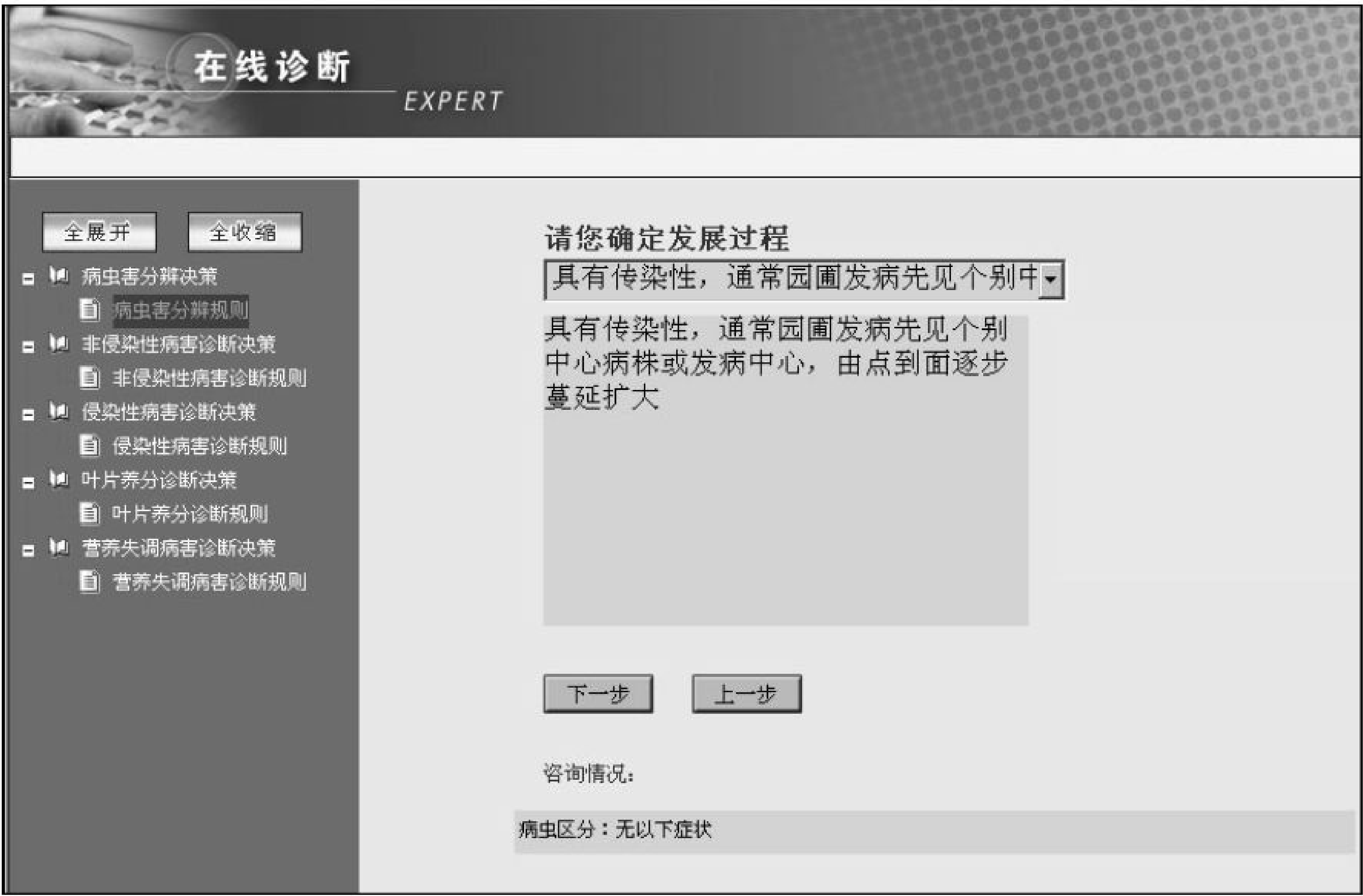


图 19-11　病虫害分辨事实表“发展过程”项



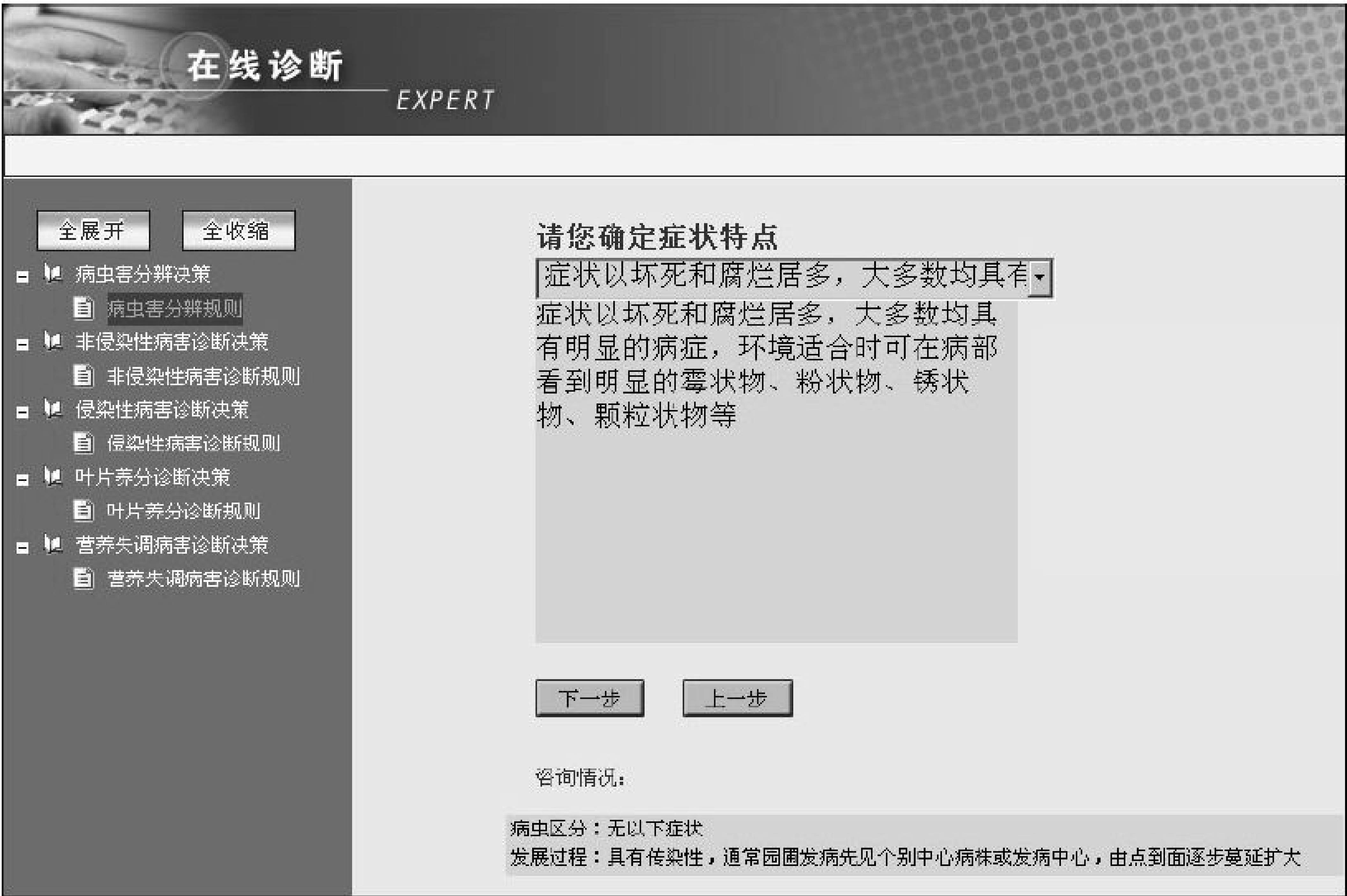


图 19-12　病虫害分辨事实表“症状特点”项



图 19-13　病虫害分辨推理显示结果



图 19-14　侵染性病害诊断事实表“部位”录入



图 19-15　侵染性病害诊断事实表“症状”录入



图 19-16　侵染性病害诊断决策结果

3. 营养失调病害诊断决策（形态诊断）

非侵染性病害诊断过程类似侵染性病害诊断过程，不再赘述。本系统的特色之一是将非侵染性病害中的营养失调病害单列出来，与前两种诊断决策不同之处在于其检索表采用自成体系的“果树矿质元素失调症诊断检索表”进行决策（图 19-17、图 19-18），最后结果给出典型图片（图 19-19）。

4. 叶片养分诊断决策

按照常量元素和微量元素进行划分（图 19-20、图 19-21），分别进行诊断，给出防治建议（图 19-22、图 19-23）。



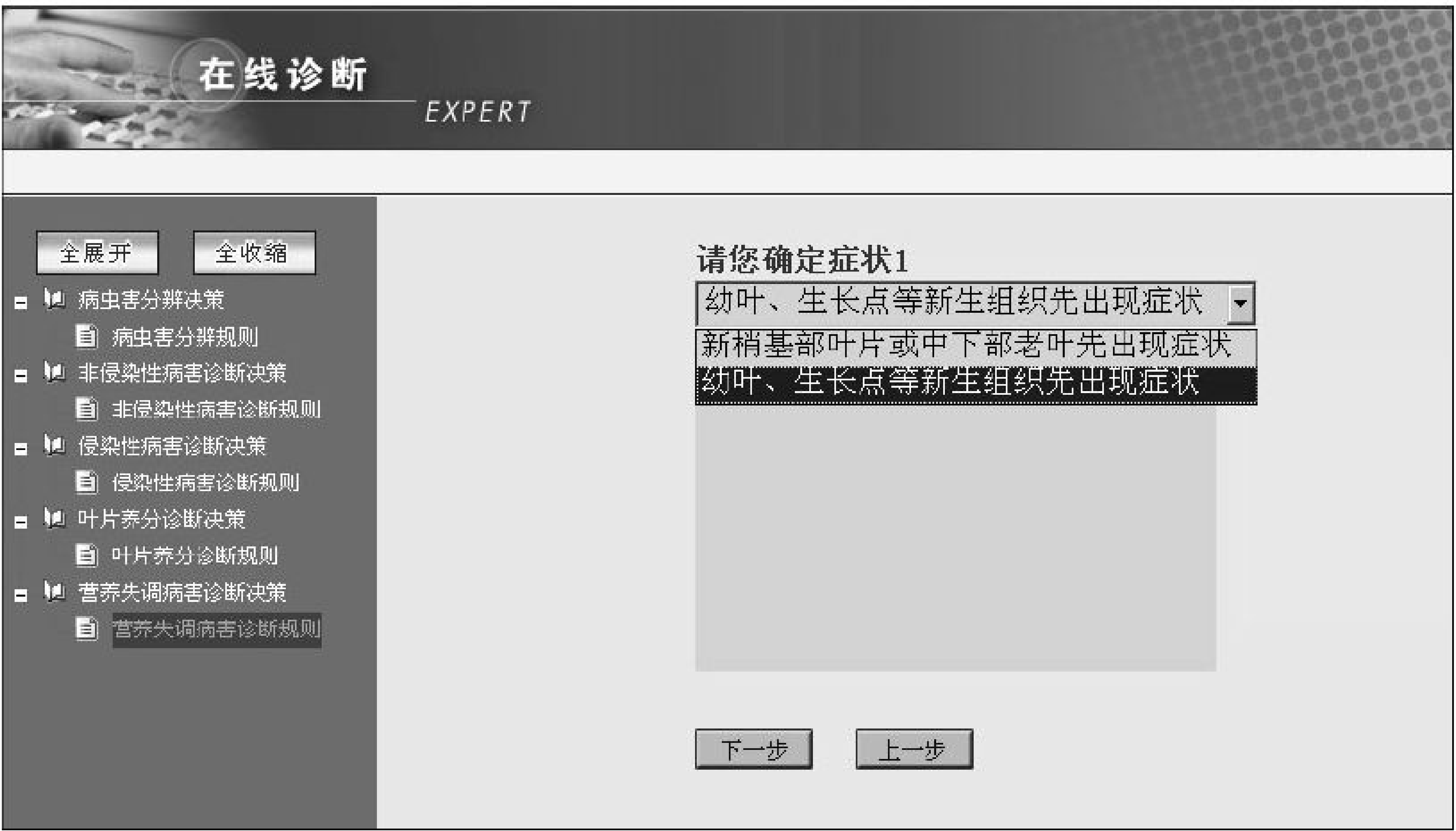


图 19-17　营养失调病害诊断事实表“部位”录入



图 19-18　营养失调病害诊断事实表“症状”录入



图 19-19　营养失调病害诊断结果



图 19-20　叶片养分诊断事实表常量元素选择



图 19-21　叶片养分诊断事实表微量元素选择



图 19-22　叶片养分诊断事实表微量元素含量输入





图 19-23　叶片养分诊断结果

19.5　系统测试

19.5.1　专家评审

由知识工程师演示系统的解决问题模式，请领域专家对其科学性和适用性进行评价，提出修改意见，再由知识工程师对模型及系统进行修改，往复进行，不断完善。

19.5.2　平台测试

通过平台的“调试”工具，可以对知识库求精，即进行逻辑语法检测，包括知识的一致性和完整性。知识的不一致性有：知识冗余、矛盾、从属和环路等。知识不完整是指知识库中的知识不完全，不能满足预先定义的约束条件，即当存在应该推出某一结论的条件时，却推不出这一结论，不能形成产生这一结论的推理链；或者虽能推出结论，但却存在错误，在知识的语法检测中检测知识是否满足 BNF 定义。还可对事实表、标准表、多媒体资料信息、安装部署信息等进行检测。

19.5.3　模拟测试

彩色照片是比较理想的纪实手段(对比黑白图和彩绘)，而且可以把病害的症状、虫害的形态、习性等主要特点表达出来，能达到看图识物的目的。用彩色照片模拟现实病虫害发生状况，结合肥城桃的病虫害具体知识，在 PAID 上实现模型，形成肥城桃病虫害诊断与防治专家系统，用其进行诊断，是一条简便易行的测试手段。通过模拟测试，该系统可以诊断出肥城桃侵染性病害 22 种，非侵染性病害 16 种，虫害 22 种。

19.6 小结

(1) 本地化、特色化、无公害化的系统开发策略

各地生态条件不同,适宜栽培的树种和品种也不同,即使相同,其管理措施也因地而异,尤其是某些果树如肥城桃,由于其品质、风味的形成具有显著的地区特异性,因此果树病虫害诊断与防治模型只有与果树专家和地方果树产业实践紧密结合,充分调查基层生产需求,按照需求进行系统分析、设计、开发,在适宜区进行周年的试验示范推广,才能开发出真正实用的模型,提高果业的经济效益和综合社会效益,为农民增收,为农村经济发展贡献力量。

(2) 模型设计遵循植物病虫害诊治原理

在本系统知识库设计中,建立病虫害分辨决策,先分辨病害或虫害,后诊断具体病害或虫害;将病害诊断分为侵染性病害诊断、非侵染性病害诊断、营养失调症诊断三部分;对营养失调症的诊断,提出形态诊断和叶片养分诊断相结合。这种设计不仅对肥城桃等果树适用,对于其他作物也具有重要的参考价值。

(3) 知识表示方式采用框架式和产生式结合的方法

本系统知识模型在开发的过程中,将知识规则构建方法固化,形成知识模型框架,利于知识库更新与维护。框架中不涉及较多计算机术语,是知识规则的浓缩,可以作为一种原型,成为知识库开发前,专家、用户和知识工程师沟通的工具,知识工程师可以根据专家和用户的意见,对于模型框架进行修改,以提高知识规则编辑的质量。由于按照统一框架进行编辑,知识规则形式有内在的统一性,当知识发生变化时,可以通过先更新框架,再对照框架更改知识规则的方法,实现知识库的更新,以适应生产的发展。框架一次编写,可多次继承使用,只要按照需要,组建合适的框架结构,填入某种园艺植物的病虫害诊断与防治知识,就可形成相应的知识库,这对其他农业专家系统知识库的构建也有重要的参考价值。

参考文献

1. 李怀方,刘凤权,郭小密.园艺植物病理学[M].北京:中国农业大学出版社,2001:7～12
2. 宗兆峰,康振生.植物病理学原理[M].北京:中国农业出版社,2002:1～19,240～244
3. 曹子刚,董桂芝.桃、梨、杏、樱桃主要病虫害及其防治[M].北京:中国林业出版社,1998
4. Guus Schreiber.知识工程和知识管理[M].北京:机械工业出版社,2003:79～106
5. 王衍安,李明,王丽辉等.果树病虫害诊断与防治专家系统知识库的构建[J].山东农业大学学报(自然科学版),2005,36 (3):475～480
6. 杨宝祝,赵春江,李爱平等.网络化、构件化农业专家系统开发平台(PAID)的研究与应用[J].高技术通讯,2002,12(3):5～9
7. 杨宝祝,赵春江,孙想.农业专家系统开发技术[J].高技术通讯,2000.10(增刊):40～44
8. 孟庆臣.ADO 与 ADO.NET.http://www.microsoft.com/china/community/program/originalarticles

第20章

重大动物疫情疫病咨询系统

本章通过分析畜禽疫病的产生、流行规律及其分布特征，确定疫病监测预警指标的获取方法和预警模型，构建和开发重大动物疫病预警与疫病咨询信息系统，为畜禽养殖业提供一个有效的动物疫病预警信息平台。

20.1 概述

随着畜禽养殖业的规模化、集约化发展，养殖密度急剧提高，环境污染日益严重，以高死亡率和突然死亡为特征的畜禽疫病暴发频率明显增加。布氏杆菌病，牛结核病，家畜血吸虫病，鸡、鸭新城疫，鸡白痢，鸡霉形体病，猪瘟及禽流感等每一次疫病的爆发都给我国畜禽养殖业带来巨大的经济损失，甚至是毁灭性的灾难。此外，畜禽携带的病毒还可能转化成新的病原直接威胁人类健康。据统计，人类历史上每次流感的大流行都与畜禽流感病毒有密切的联系。[1]

传染源、传播途径以及易感群体是畜禽疫病流行的主要条件，控制疫病的最有效方法首先是迅速切断传染源，其次是有效控制畜禽疫病的传播途径，最后是保护易感群体。[2]因此，确定有效的监测指标和敏感快速的监测方法，迅速发现传染源，预测疫情并及早发现疫情，把疫情控制在最小范围内，防止传染源进入市场流通渠道，是实施有效疫病防控的基础。[3]我国畜禽养殖疫病在动物医学基础上已有深入研究，具备一定条件，并且对预防、控制畜禽疫病，保护生产，有极其重要的经济价值和社会效益，但是针对畜禽疫病的监测、预警和防控信息体系的建设和发展还比较缓慢。

本系统以重大动物疫病中的禽流感、口蹄疫等典型疾病为研究对象，开发一套针对我国畜禽养殖业疫情疫病的咨询系统，为减少畜禽疫病带来的各种损失，实现畜禽疫病的有效预警提供可靠的理论依据。

20.2 系统分析与设计

20.2.1 系统分析

1. 系统用户分析

本系统的用户主要有两类：养殖场普通用户和畜禽养殖技术人员。对于养殖场普通用户来说，系统的主要用途是对疫情信息、历史信息的咨询和预警功能，将疫情信息、预警警级和预警预案以网页或短信等方式输出。对于畜禽养殖技术人员来说，使用者需要掌握简单的电脑

知识，从而对后台的各种数据信息进行操作处理，实现预警指标以及疫病预警等信息的增、删、改、查等。

2. 功能需求分析

系统功能需求主要包括如下四个方面：

(1) 用户信息、用户权限、预警信息、畜禽养殖信息的管理；

(2) 预警指标权重、预警知识库、预警模型的管理，预警指标权重运算和一致性分析；

(3) 重大动物疫病知识咨询、疫病预防知识咨询、养殖知识咨询；

(4) 历史数据查询、疫病预警功能。

20.2.2 系统设计

1. 功能模块设计

重大动物疫病预警系统的主要功能包括：知识咨询模块、疫病预警模块和系统维护模块，每个功能模块又由若干个子模块组成。

知识咨询模块包括动物养殖知识、重大动物疫病知识和疾病预防知识等，用户可以通过浏览器浏览相关知识，也可以根据需求通过关键字对各类知识进行查询、检索。疫病预警模块包括动物疫病预警(对根据用户的输入信息得出重大动物疫病的预警警级和预警预案)、预警知识库查询(预警警级与预警输入因子之间关系的知识查询)、预警历史记录查询以及动物疫病疫情记录的查询；系统维护模块是实现对用户信息、预警记录、预警指标、疫病知识等相关信息的添加、修改与删除等操作(见图 20-1)。

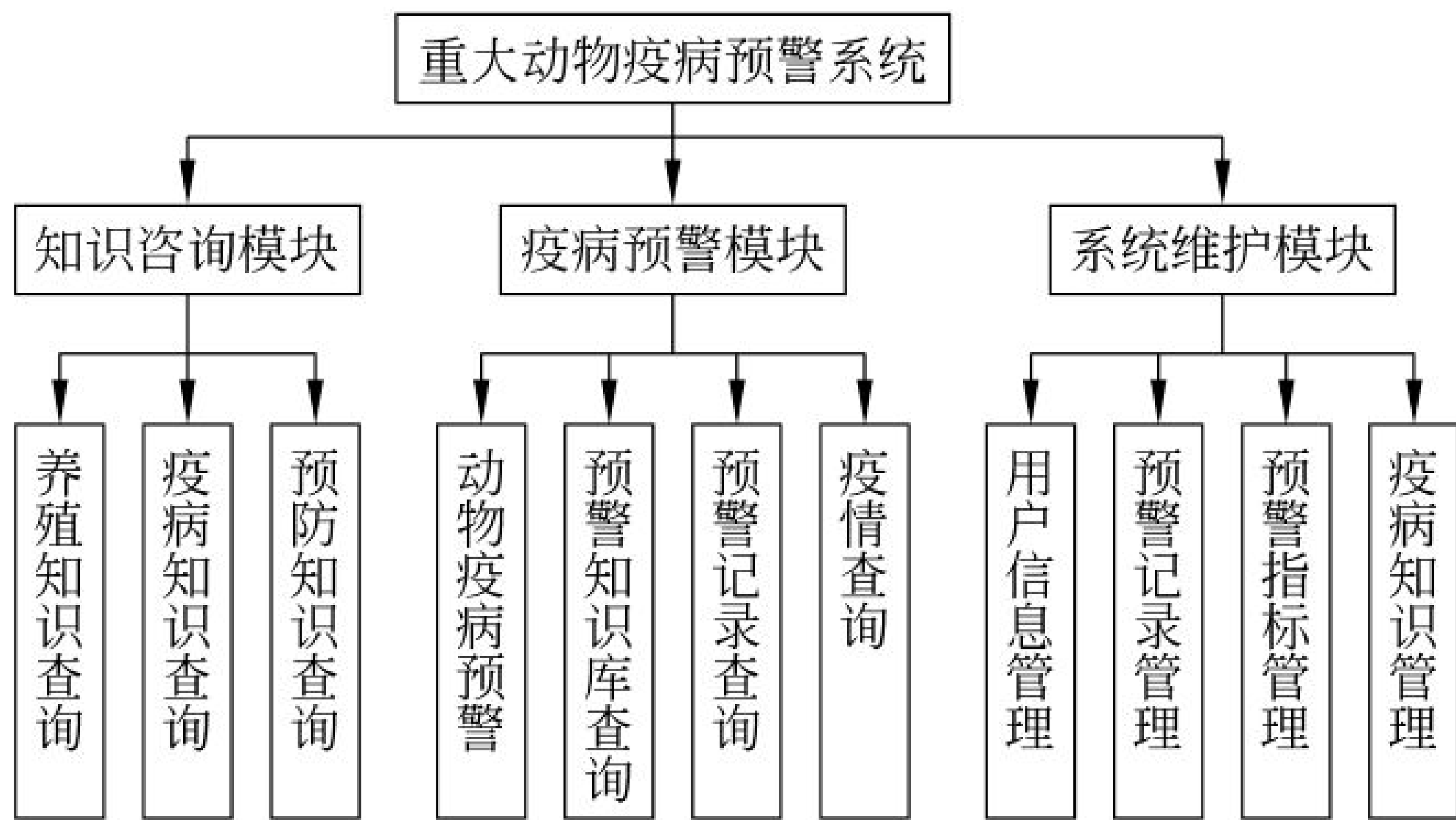


图 20-1　预警系统功能模块

2. 预警模型设计

预警模型建立的基本步骤是：首先，通过文献分析和与专家交流的方式确定影响疫情发生的预警指标；其次，结合德尔菲专家调查法和层次分析法确定每个预警指标的权重大小；最后，建立预警警级与预警预案的知识库(预警指标及其权重的确定可参见本章附录)。当用户输入相应的疫情信息时，系统可以根据每一指标的权重和单个指标级别的对应关系得出综合警级大小的数值，然后通过查询所建的知识库得出预警警级和预警预案信息。下面以禽流感预警模型设计为例，建立疫病预警模型。



(1) 预警指标选取

以禽流感疫病预警为例,选取影响疫情发生的预警指标,包括饲养模式、禽类本身、外来因素、疫情因素、禽类流通等方面,具体预警指标如图 20-2 所示。

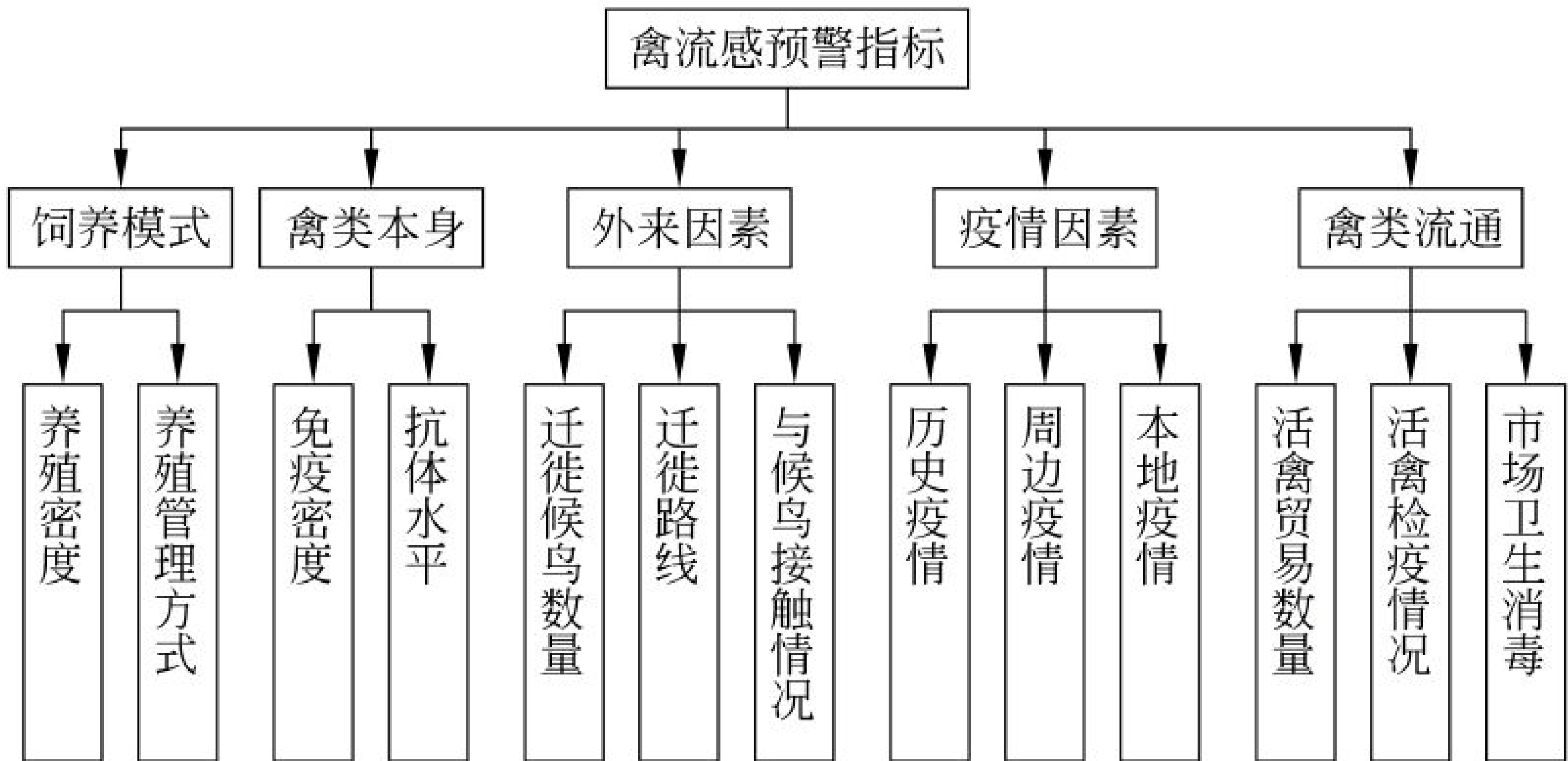


图 20-2　禽流感预警指标

(2) 预警指标警级区间

预警指标警级区间的确定是采用专家调查法。为了满足预警评价的要求,对每一个影响因子,需要根据专家调查的方法,综合多位动物疫病专家的意见,来确定每个预警指标的边界点。其中定性的预警指标如养殖管理方式、与候鸟接触情况等采用综合多个专家对不同警级信息的描述方法确定单个因素预警警级的描述信息;而定量的预警指标如养殖密度、抗体水平等,则采用下式确定预警指标的轻警、中警、重警的预警指标警级边界值。

$$V = \frac{\sum E(w_i) \times a_i}{N}$$

其中 $E(w_i)$为每个专家所占的权重大小,它的大小主要是根据专家的工作经验、年龄、学历等参数来确定的,a_i 为每个专家对此边界点设定的值,N 为专家数量。

(3) 指标权重的确定

预警指标权重的确定采用层次分析法,层次分析法主要是将所分析的问题层次化,将问题分解成不同的组成因素,再按照因素间的相互关系及隶属关系,将其聚集组合,形成一个统分结合的分析结构模型。以禽流感预警指标为例,首先把禽流感预警指标分为饲养模式、禽类本身、外来因素、疫情因素、禽类流通等,然后对每一类别进行再次细分。

$A=\{C1,C2,C3,C4,C5,C6,C7,C8,C9,C10,C11,C12,C13\}$,式中,$A$ 为问题全集,对应禽流感预警警级大小,$C1 \sim C13$ 为 13 个预警指标。上述 13 个预警指标归并为五个方面的问题(饲养模式、禽类本身、外来因素、疫情因素和禽类流通)。

$$\mathrm{COV}(A) = \left\{ \frac{B1(\text{饲养模式})}{C1,C2}, \frac{B2(\text{禽类本身})}{C3,C4}, \frac{B3(\text{外来因素})}{C5,C6,C7}, \frac{B4(\text{疫情因素})}{C8,C9,C10}, \frac{B5(\text{禽类流通})}{C11,C12,C13} \right\}$$

根据专家调查法得出 B 类之前的对比矩阵、C 类之间的矩阵,计算 $B1 \sim B5$,$C1 \sim C13$ 的权重值。

20.2.3　预警预案设计

预警预案是指面对可能造成重大人员伤亡、财产损失、生态环境破坏和严重社会危害的重

大动物疫病疫情发生，采取各种应急管理、计划等，以消除或减小疫情所带来的损失。以禽流感为例，对禽流感疫情所采取的措施主要由以下几部分组成。

(1) 管理传染源

加强禽类疫情监测；对受感染动物应立即销毁，对疫源地进行封锁，彻底消毒；患者应隔离治疗，转运时应戴口罩。尽早发现禽流感病禽和病人；向卫生防疫部门报告禽流感病禽和病人；对病人要至少隔离至退热后两天，病禽要封闭或封锁；要尽快治疗病人，杀灭病禽。对病人要进行综合性有效治疗，在病鸡场周围一定范围内的病禽就地杀灭、深埋。[4]

(2) 切断传播途径

接触患者或患者分泌物后应洗手；处理患者血液或分泌物时应戴手套；被患者血液或分泌物污染的医疗器械应消毒；发生疫情时，应尽量减少与禽类接触，接触禽类时应戴上手套和口罩，穿上防护衣。禽流感病人、接触者(如医护人员和饲养人员)必须戴口罩；病房、养鸡场和居室加强通风换气；远离易感场所：少去或不去人群密集的场所或养鸡场，去时戴口罩；消毒：对病房和养鸡场的空气消毒；对病人和病禽的分泌物与排泄物消毒；对被病毒污染的物体表面消毒。[4]

(3) 减少易感人群或高危人群

60 岁以上老人、儿童、小学生、免疫力低下者、慢性病患者是流感的好发人群，对这些人要注意御寒，加强户外锻炼，增强抵抗力，接种流感疫苗；远离家禽的分泌物，尽量避免触摸活的鸡、鸭等家禽及鸟类；保持室内空气流通，应每天开窗换气两次，每次至少 10 分钟，或使用抽气扇保持空气流通；保持地面、天花板、家具及墙壁清洁，确保排水道去水顺畅；吃禽肉要煮熟、煮透。注意多摄入一些富含维生素 C 等增强免疫力的食物和药物；适当进行体育锻炼，以增加机体对病毒的抵抗能力；勤洗手。[5]

通过上述对预警措施的分析，需要建立针对不同警级而采取不同预警措施的预警预案知识库。预警预案主要是针对轻警、中警、重警等不同级别的警级，对捕杀半径、禽类交易人员的监测控制等各种措施实行不同程度的执行。

表 20-1 预警预案表

编号	预警预案	预 案 表 述	轻警	中警	重警	巨警
1	捕杀半径	对疫区内的禽类实施杀灭、深埋				
2	禽类交易	对疫区和周边一定范围内的禽类交易场所采取不同程度的控制措施				
3	人员的监测	对疫区和周边一定范围内的人口和禽流感密切接触者进行身体体征的监测或隔离处理				
4	疫区封锁	对疫区和周边一定范围实施不同程度的封锁措施				
5	预防预警公告	向公众发布多种信息，包括预防方法、当前疫情、疫情变化趋势、安全知识宣传等公告				
6	养殖地管理	对疫区和周边一定范围内的禽类养殖地实施有行预警、处理等管理措施				
7	非疫区控制	对非疫区进行不同程度预防监控措施				
8	疫病的监测	针对不同的疫情实施不同程度的监测力度				
9	消毒	消毒处理的强度和范围				



禽流感预警流程如图 20-3 所示。

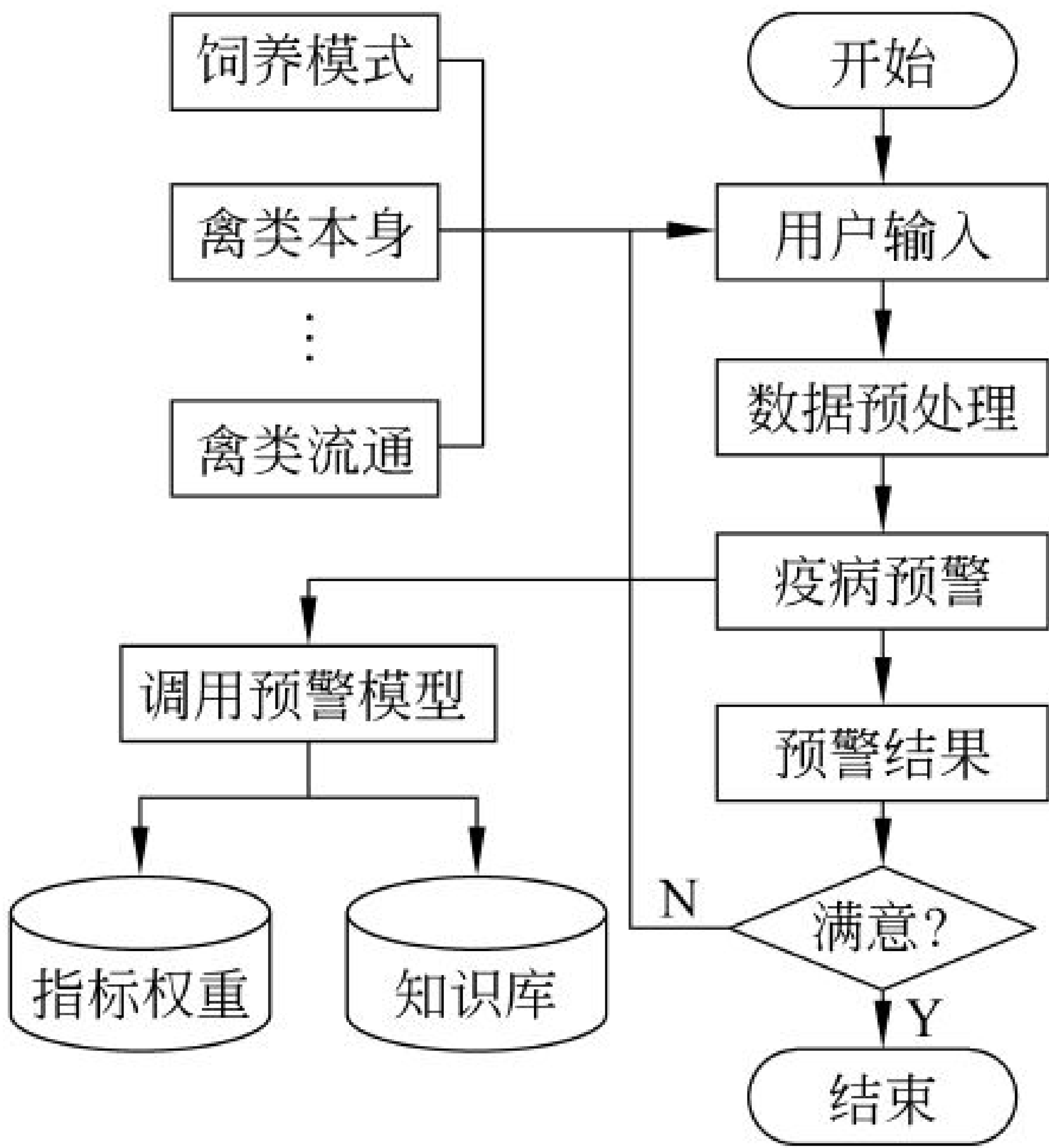


图 20-3 禽流感预警流程图

20.2.4 数据库设计

重大动物疫病预警系统数据库主要包括用户信息表、预警记录表、预警知识库表、预警指标表。其中预警记录表用于记录历史疫情数据，预警指标采用指标集方法以字符串的形式存储。在养殖、疾病知识表中，由于需要对各种知识进行分级表示，实现树形结构展示，因此增加了父节点编号这一字段，用于标识该内容的层次信息。考虑到数据存取的效率和便捷问题，用二维表形式存储各预警指标，相互对比的分值信息则存储在 excel 文件（如表 20-2～表 20-6 所示）。

表 20-2 用户信息表

字 段	类 型	允许空	主/外键	备 注
id	int	no	PK	编号
username	int	no		用户名
password	Varchar	no		密码
authority	int	no		权限
name	Varchar	yes		姓名
sex	Varchar	yes		性别
age	Varchar	yes		年龄
telephone	Varchar	yes		联系电话
email	Varchar	yes		电子邮件
address	Varchar	yes		通信地址

表 20-3 预警记录表

字 段	类 型	允许空	主/外键	备 注
id	int	no	PK	编号
warningid	int	no	FK	预警警级大小
userid	int	no	FK	记录人编号
diseaseaddress	Varchar	yes		地点
time	datetime	yes		预警时间

表 20-4 预警知识库表

字 段	类 型	允许空	主/外键	备 注
id	int	no	PK	编号
indexset	int	no		指标记录集
warninglevel	Varchar	no		预警警级
measure	int	no		预警预案
prevent	Varchar	yes		预防措施

表 20-5　预警指标表

字　段	类　型	允许空	主/外键	备　注
id	int	no	PK	编号
indexname	int	no		指标名称
indextype	Varchar	no		指标类型
indexweight	int	no		指标权重
firstlevel1	Varchar	no		轻警上限值
firstlevel2	Varchar	no		轻警下限值
secondlevel1	Varchar	no		中警上限值
secondlevel2	Varchar	no		中警下限值
content	Varchar	yes		指标描述

表 20-6　养殖、疾病知识表

字　段	类　型	允许空	主/外键	备　注
id	int	no	PK	编号
name	int	no		知识标题
type	Varchar	no		知识类型
content	int	no		知识内容
pid	Varchar	no		父节点编号
remarks	Varchar	no		备注信息

20.3　系统实现

20.3.1　系统开发环境和相关实现技术

重大动物疫情疫病预警系统以 Eclipse 为开发工具，在 Windows XP＋JDK 1.5.0＋Tomcat 5.0＋SQL Server2000 环境下进行开发、编译及调试，如图 20-4 所示。采用基于 Struts 的 MVC 开发模式，Struts 是目前比较流行的基于 MVC 的框架，在总体上实现了对业务逻辑层、显示层和控制层的分离，同时也提高了应用软件的可扩展性和重用性。

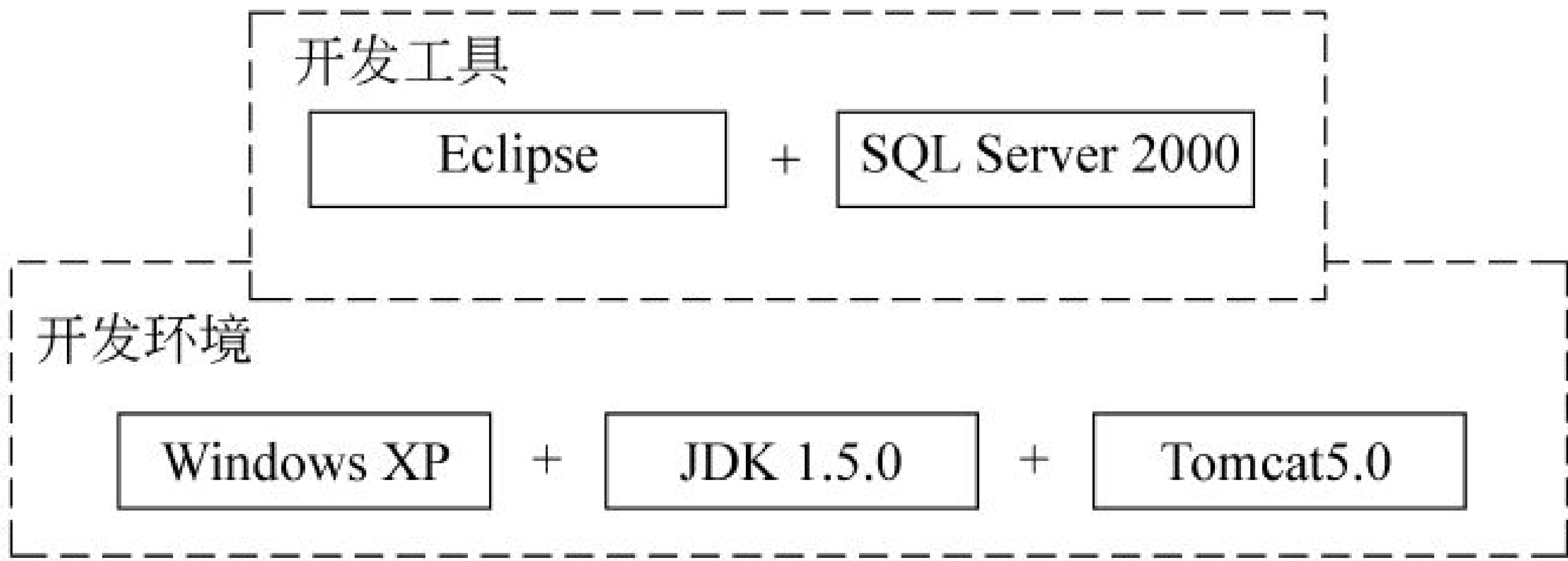


图 20-4　系统开发的软件环境

20.3.2　系统运行界面

系统的主要功能实现界面如图 20-5～图 20-7 所示。





图 20-5　禽流感预警指标

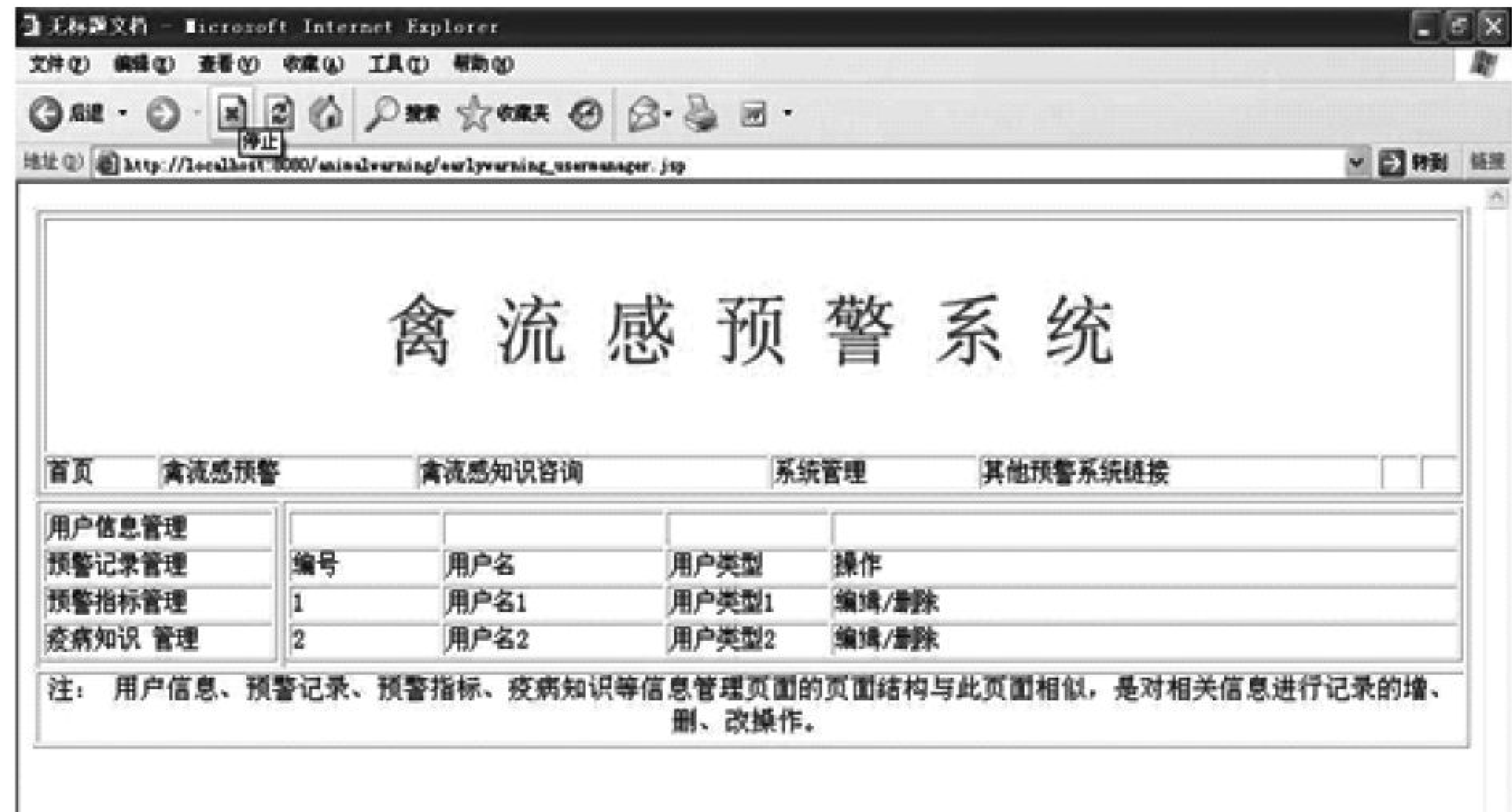


图 20-6　禽流感预警指标



图 20-7　禽流感预警指标

20.4 小结

本章从系统需求出发，对系统的结构、功能模块、流程和数据库进行了详细设计，配置了系统的软件平台，并采用基于 Struts 的 MVC 开发模式，以 Java 为开发语言，以重大动物疫病中的禽流感、口蹄疫等典型疾病为研究对象，开发出一套针对我国畜禽养殖业疫情疫病的咨询系统。

参考文献

1. 李道亮. 重大动物疫病预警与防控技术体系研究[C]. 中国畜牧兽医学会信息技术分会 2007 年学术研讨会，2007
2. 宋建德，郑雪光，滕翔雁等. 动物疫病预警体系建设初探[J]. 中国动物检疫，2005，22(3)：10～12
3. 郭艳秀. 建立天西林区野生动物疫源疫病监测预警体系的思考[J]. 中国林业经济，2007(2)：57～58
4. 肖剑. 温州市重大动物疫病预警预报监测体系建设[J]. 中国禽业导刊，2009(4)：18
5. 杨丁. 动物疫病早期预警体系的构建. 中国牧业通讯，2005(3)：31～34

附录

本附录提供的表格可用于确定预警指标、预警区间及其权重。

附表 1　预警指标及其预警区间

编号	预警指标	指标类型	轻　　警	中　　警	重　　警
1	养殖密度	饲养模式	＜1000 只/平方千米	1000～5000 只/平方千米	≥5000 只/平方千米
2	管理方式	饲养模式	该地区集约化、规模化养殖所占比重大，养殖区域集中，养殖场舍隔离、卫生与消毒状况良好	该地区以小规模养禽场为主，集约化、规模化水平较低，卫生与消毒状况较好	该地区以农户养殖为主，散养、混养形式居多，养殖地区分散，生产环境不能封闭隔离，饲养密度高，卫生与消毒状况极差
3	免疫密度	禽类本身	80％以上	30％～80％	低于 30％
4	抗体水平	禽类本身	HI 抗体水平≥6log2，可持续 6～10 个月	4log2≤HI 抗体水平＜6log2，可持续时间 4～5 个月	HI 抗体水平＜4log2，可持续时间 2～3 个月
5	迁徙候鸟数量	外来因素	迁徙候鸟数量很少	经该省的迁徙候鸟数量适中	经该省的迁徙候鸟数量多
6	迁徙路线	外来因素	迁来时不经过禽流感疫区	少量候鸟迁来时经过禽流感疫区	大部分候鸟迁徙时经过禽流感疫区
7	与候鸟接触	外来因素	家禽封闭式养殖，家禽与候鸟直接或间接接触很少，自然保护区、湿地和水系较少，与候鸟无共用的水源	大部分封闭式养殖，有少量自然保护区、湿地分布	露天养禽数量较多，多有候鸟在养禽场栖息，有重要的大型自然保护区、湿地分布并水网密集，且与候鸟共用的水源有多处



续表

编号	预警指标	指标类型	轻　　警	中　　警	重　　警
8	历史疫情	疫情因素	未发生疫情	0<疫点数量<5 或发生过疑似疫情	以前发生过禽流感疫情，疫点数量≥5 或发病禽只数量≥5000 只
9	周边疫情	疫情因素	距该地 500 千米内未发生疫情	距该地 500 千米内有疫情，且疫点数=1 或发生过疑似疫情	距该地 500 千米内有疫情，且疫点数量≥2 或发病禽只数≥2000 只
10	贸易数量	禽类流通	少量、小规模的活禽交易	中等规模的贸易	大量、频密引进活禽或向其他远距离地区销售大量活禽
11	检疫情况	禽类流通	严格的检疫环节，严禁病禽进入市场流通	较好的检疫程度，但有漏检现象	无检疫环节
12	卫生消毒	禽类流通	卫生状况良好，及时清理家禽粪便并进行消毒处理，对笼具和运输车辆严格消毒	卫生状况较好，及时清理家禽粪便，对笼具和运输车辆消毒较好	卫生状况极差，不能及时清理家禽粪便，对笼具和运输车辆消毒不严格

附表 2　预警指标比较原则

标度 a_{ij}	定　　义
1	因素 i 与因素 j 相同重要
3	因素 i 比因素 j 稍重要
5	因素 i 比因素 j 较重要
7	因素 i 比因素 j 非常重要
9	因素 i 比因素 j 绝对重要
2，4，6，8，	因素 i 与因素 j 的重要性的比较值介于上述两个相邻等级之间
倒数 1，$\frac{1}{2}$，$\frac{1}{3}$，$\frac{1}{4}$，$\frac{1}{5}$，$\frac{1}{6}$，$\frac{1}{7}$，$\frac{1}{8}$，$\frac{1}{9}$	因素 j 与因素 i 比较得到判断值为 a_{ij} 的相反数，$a_{ji}=\frac{1}{a_{ij}}$，$a_{ii}=1$

以下附表 3～附表 6 要依据附表 2 给出的原则填写。

附表 3　不同预警类别重要性比较

	饲养模式	禽类本身	外来因素	疫情因素	禽类流通
饲养模式					
禽类本身					
外来因素					
疫情因素					
禽类流通					

附表 4　饲养模式预警指标比较

	养殖密度	管理方式		免疫密度	抗体水平
养殖密度			免疫密度		
管理方式			抗体水平		

附表 5　外来因素预警指标比较

	迁徙候鸟数量	迁徙路线	与候鸟接触
迁徙候鸟数量			
迁徙路线			
与候鸟接触			

附表 6　疫情因素预警指标比较

	历史疫情	周边疫情	本地疫情
历史疫情			
周边疫情			
本地疫情			

第21章

基于Web的牙鲆疾病预警系统

本章在对牙鲆疾病预警模型构建的基础上，对系统的结构、功能模块、工作流程和知识库进行了详细设计，开发了基于Web的牙鲆疾病预警系统。该系统的核心工作是对主要功能模块的程序设计，最后实现水环境状态预警、趋势预警、非水环境预警和症状预警模块的集成。

21.1 系统分析

21.1.1 用户需求分析

牙鲆疾病预警系统用户主要有三类：第一类是牙鲆养殖用户，系统为用户提供疾病预警和信息查询等功能；第二类是养殖技术人员和牙鲆疾病专家，除了进行牙鲆疾病预警，他们还对系统中的疾病案例库和知识库进行维护；第三类是系统管理员，负责对系统进行维护，包括用户管理、信息维护等操作。[1]

21.1.2 功能需求分析

系统实现水环境预警、非水环境因素预警、疾病诊断和症状预警等功能，并以网页、手机短消息等多种形式通知用户预警警级和预警预案；实现牙鲆疾病案例、知识、历史预警记录查询功能，为用户决策提供辅助支持；实现系统各类信息维护功能。

21.2 系统设计

21.2.1 系统结构设计

基于Web的牙鲆疾病预警系统可按系统的结构层次划分为：数据接收层，业务处理层和结果显示层，如图21-1所示。下面介绍各结构层。

(1) 数据接收层是牙鲆疾病系统中的信息获取部分，由三部分组成：传感器接收的数据，手工检测得到的数据和病症信息。

(2) 业务处理层通过对三类接收信息进行综合分析和处理，得出预警警级大小和预警预案。本层包括水环境预警子系统，非水环境预警子系统和症状预警子系统。

(3) 结果显示层将业务处理层处理后的结果以短信或者网页形式显示给用户。

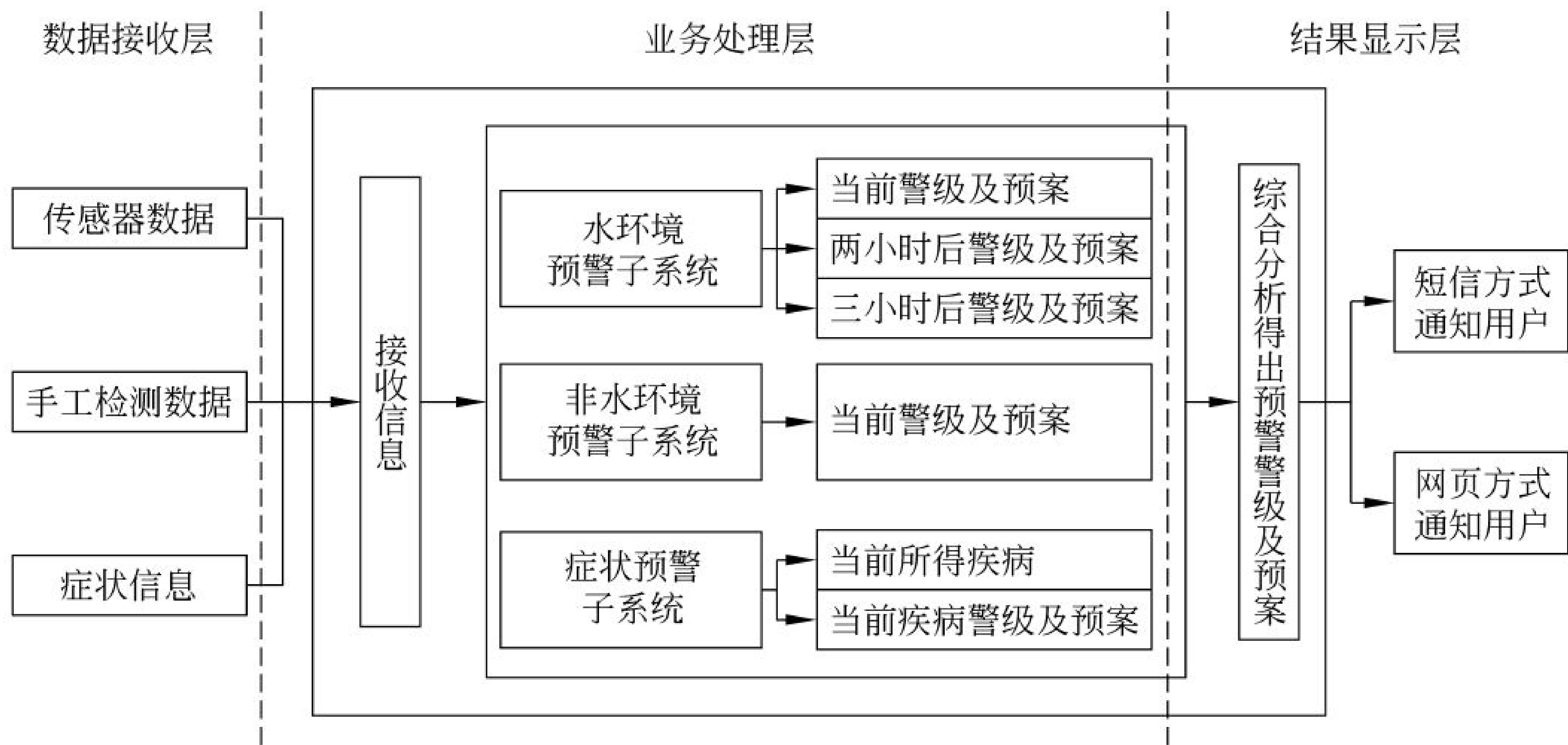


图 21-1　系统结构

21.2.2　功能模块设计

本系统的功能模块包括：鱼病预警模块、知识浏览模块和系统维护模块。每个功能模块由若干个子模块组成，如图 21-2 所示。

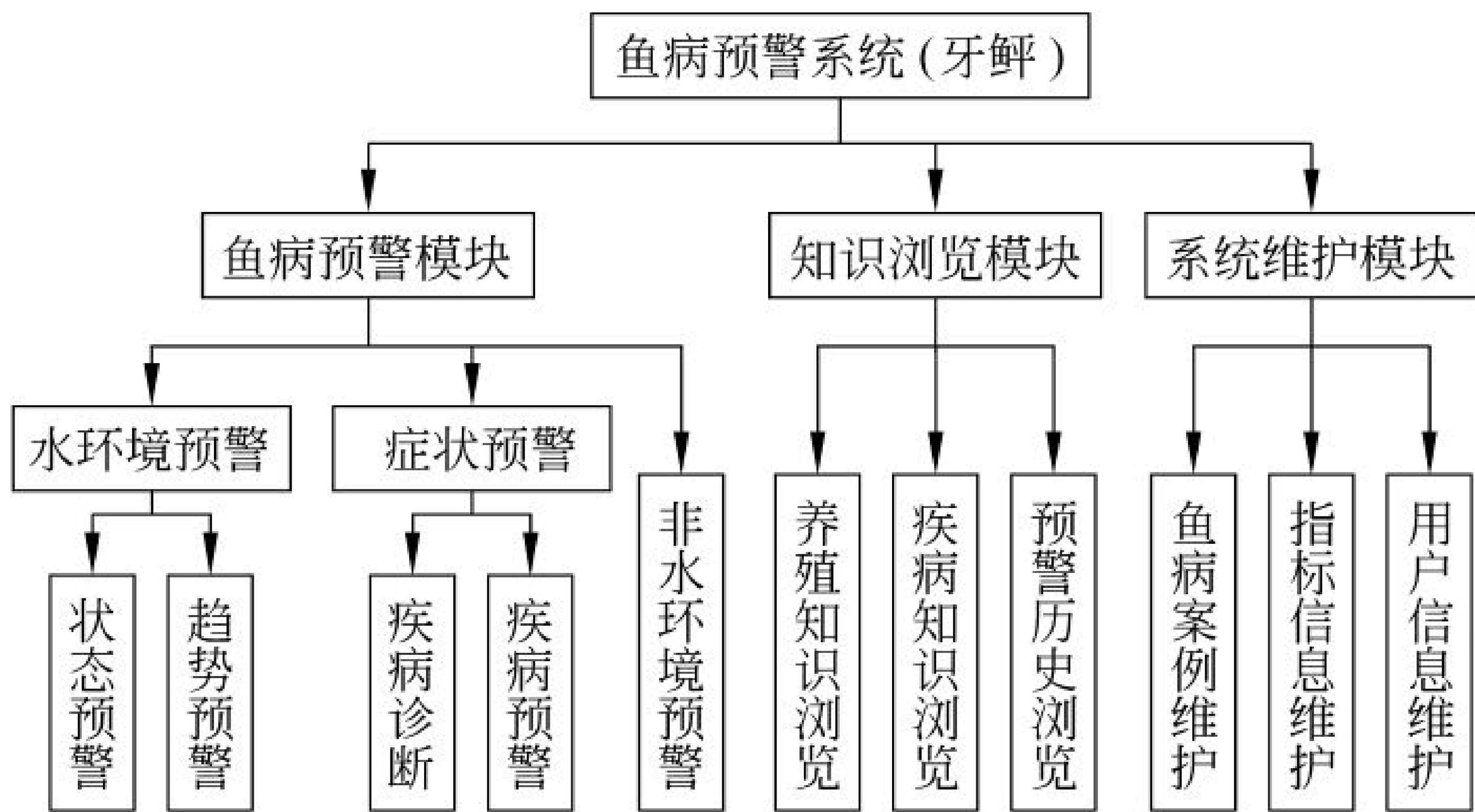


图 21-2　系统功能模块

（1）鱼病预警模块

鱼病预警模块分为水环境预警、非水环境预警和症状预警三部分，[2]其中水环境预警包括对当前水质的评价预警，对未来水质预测后的评价预警，即水环境状态预警和趋势预警，下面分别对这些模块进行介绍。

水环境预警：利用专家调查的方法，确定集约化牙鲆养殖的主要影响因素为溶氧、水温、盐度、氨氮、pH、COD、硝酸根离子含量七个水质参数为准的预测预警。就牙鲆养殖而言，对水环境有特定要求，因此，需要根据专家调查的方法，综合多个水产养殖专家的意见，来确定每个水质参数的无警、中警、重警的边界点，进而确定每一个警级的警级区间。然后将参数的警级区间进行排列组合，参考多个专家的意见确定每一种情况的警级大小和预警预案。

水环境趋势预警：利用 BP 神经网络与遗传算法相结合的方法，根据当前水环境各个参数值，预测两个小时或三个小时后水环境的各个参数值，然后再利用状态预警的方法得出两个小



时和三个小时后的警级大小和预警预案。

非水环境预警：通过对寄生虫含量、饵料质量、鱼体损伤、养殖密度的评价，确定当前的警级大小和预警预案。[3] 其中鱼体损伤根据无损伤、轻损伤和重损伤所占百分比来确定此因素的警级区间，而其他因素则同样按专家调查方法确定每个因素的警级区间，非水环境预警主要是对单因子进行评价，若某一个因素超过确定的警限就输出相应的预警预案。

症状预警：包括疾病诊断和疾病预警两部分。首先根据专家知识得出不同疾病在不同发病率的警级大小。[4] 当用户输入症状时，对牙鲆疾病进行诊断，得出疾病诊断结果，然后再根据用户所输入的有此症状的牙鲆的发病率来确定症状预警警级的大小。其中牙鲆疾病的诊断采用基于知识与基于案例相结合的方法。

（2）知识浏览模块

知识浏览模块包括牙鲆养殖知识、牙鲆疾病和疾病预防知识等，养鱼户可以通过浏览器浏览相关知识，也可以通过关键字对各类知识进行搜索、查询。

（3）系统维护模块

系统维护模块实现对牙鲆疾病案例、预警指标、用户信息的维护，包括对疾病案例和预警指标等数据的添加、修改与删除等操作。

21.2.3　系统流程设计

牙鲆疾病预警系统流程示意图如图 21-3 所示。

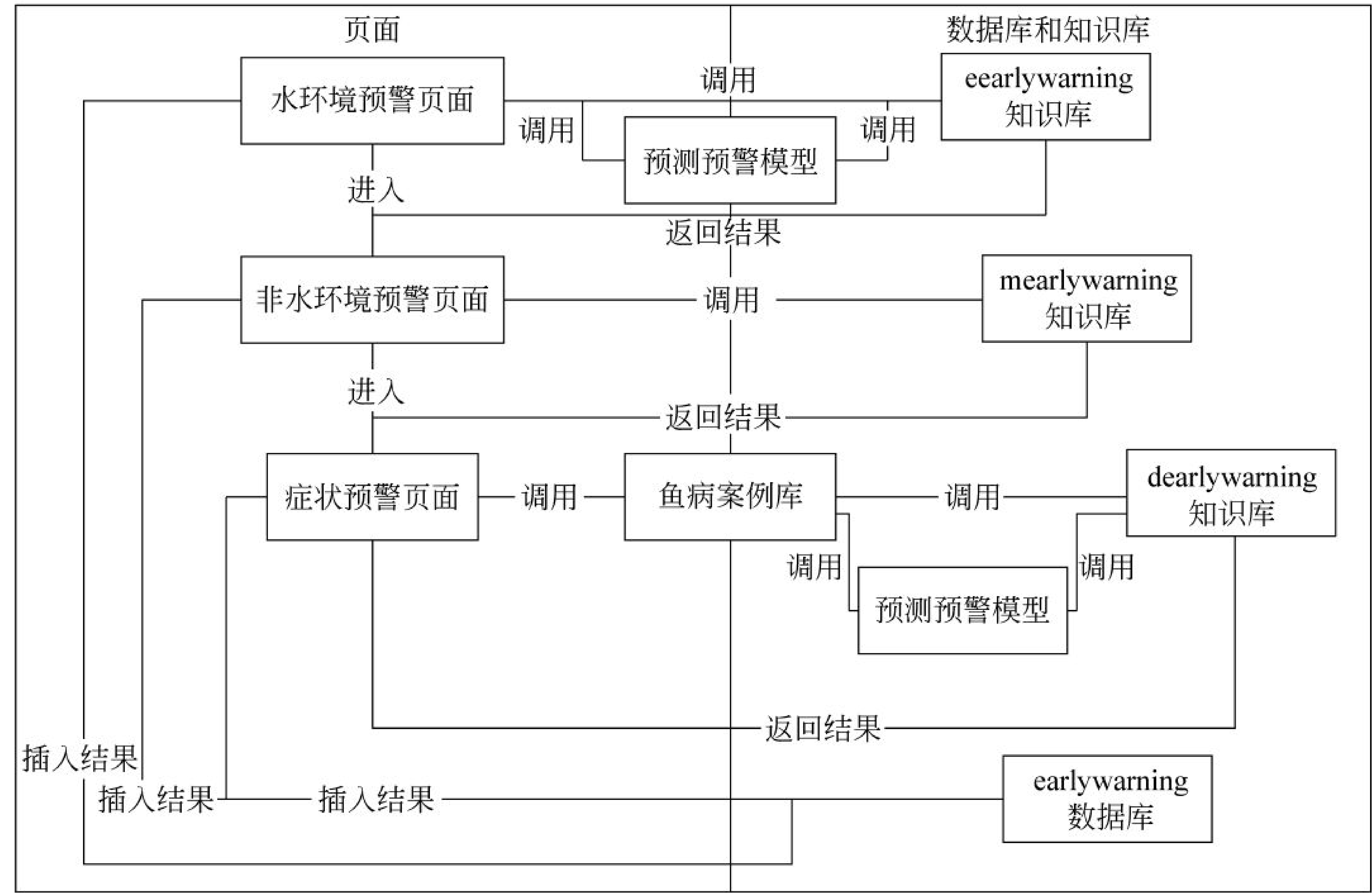


图 21-3　系统流程示意图

21.2.4　数据库设计

本系统数据库包括水环境预警表、非水环境预警表、症状预警表、警情信息发布表、症状

表、案例表。其中水环境预警表、非水环境预警表、症状预警表分别用于确定每一种预警的警级大小,警情信息发布表用于对所有预警信息的综合发布,症状表、案例表用于对牙鲆疾病进行诊断。其中,案例表中案例集合字段中存储的是每一案例对应的症状编号组合,疾病名称可以根据疾病编号从鱼病知识表中查询得到。数据库结构如图 21-4 所示。

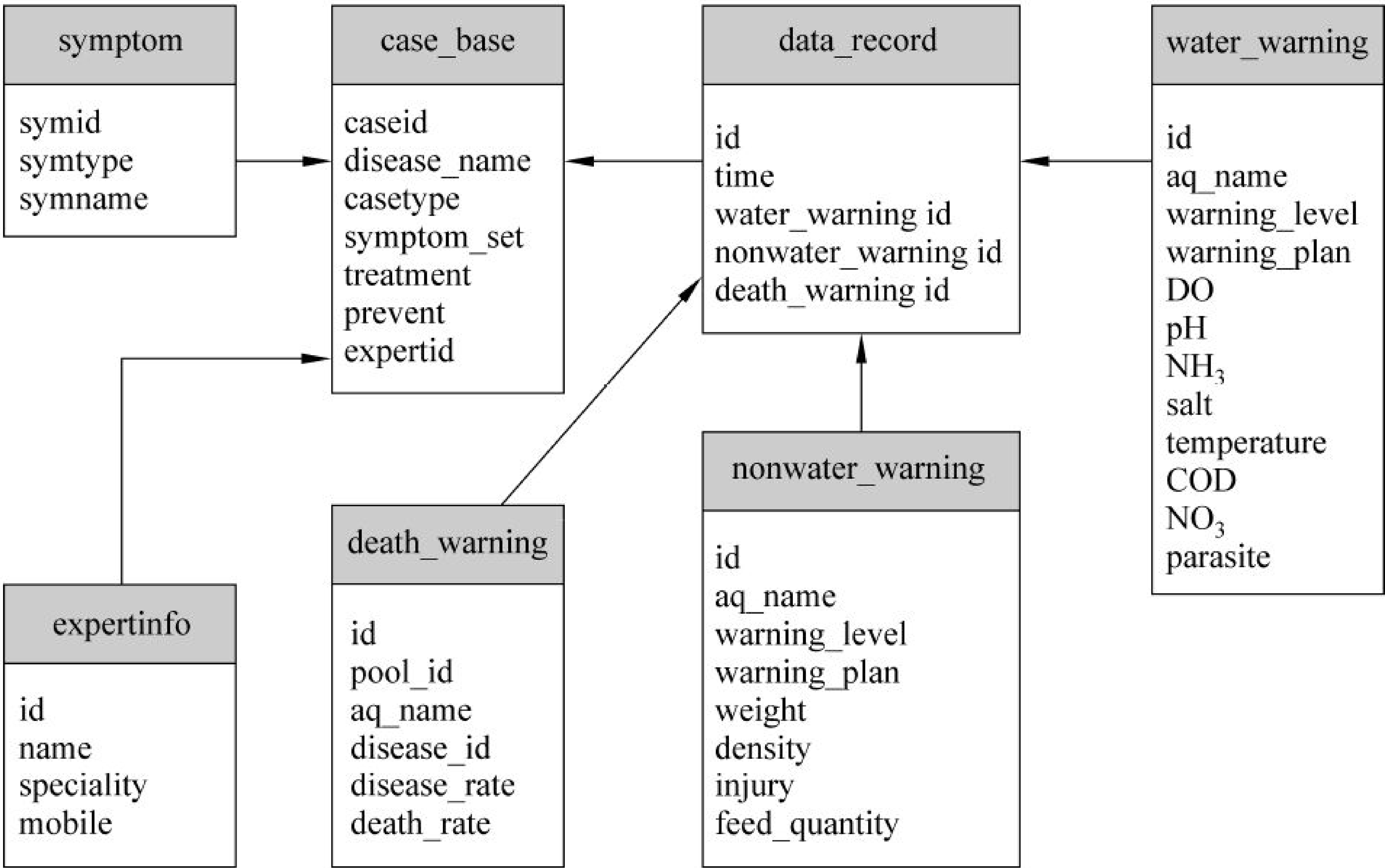


图 21-4　数据库结构图

图 21-4 中的数据表设计如下:

(1) 水环境预警表(编号,水产品名,警级大小,预警预案,溶氧,pH,氨氮,盐度,水温,需氧量,硝酸盐含量,寄生虫含量),即 water_warning(id,aq_name,warning_level, warning_plan,DO,pH,NH_3,salt,temperature,COD,NO_3,parasite)。

(2) 非水环境预警表(编号,水产品名,警级大小,预警预案,水产品规格,养殖密度,损伤程度,饵料质量),即 nonwater_warning(id, aq_name,warning_level, warning_plan, weight, density, injury, feed_quality)。

(3) 死亡预警表(编号,水池编号,水产品名,疾病编号,发病率,死亡率),即 death_warning(id, pool_id, aq_name, disease_id, disease_rate, death_rate)。

(4) 数据记录表(编号,时间,水环境预警编号,非水环境预警编号,死亡预警编号),即 data_record (id, time, water_warning id, nonwater_warning id, death_warning id)。

(5) 鱼病症状表(症状编号,症状类型,症状名称),即 symptom(symid, symtype, symname)。

(6) 鱼病案例表(案例编号,疾病名称,案例类型,症状集,治疗措施,预防办法,专家编号),即 case_base(case_id, disease_name, casetype, symptom_set,treatment, prevent, expertid)。

(7) 鱼病专家信息表(专家编号,姓名,专业特长,联系方式),即 expertinfor(id,name, speciality,mobile)。



21.3 系统实现

21.3.1 系统开发软件环境和软件技术

本系统以 Eclipse＋SQL Server2000 为开发工具，在 Windows XP＋Tomcat 5.0 环境下进行开发、编译及调试。采用基于 Struts 的 MVC 开发模式，建立了基于 Web 的 B/S 体系结构的集约化牙鲆养殖疾病预警系统。Struts 是目前比较流行的基于 MVC 的框架，在总体上实现了对业务逻辑层、显示层和控制层的分离，同时也提高了应用软件的可扩展性和重用性。[5,6]下面对开发系统所用到的软件环境和技术进行简要介绍。

SQL Server 2000 是 Microsoft 公司开发和推广的大型关系数据管理系统，具有非常强大的关系数据库创建、开发设计以及管理功能。它提供对 Web 技术的支持，适用于 B/S 模式下的关系型数据库管理系统。它可以在一个数据库文件中管理所有的数据信息，并为用户提供了强大的数据处理功能、简单的图形化管理工具、丰富的编程接口工具，并支持多线程操作和分布式事务管理。可以帮助用户组织和共享数据库信息，使用户方便地得到所需数据。

Tomcat 5.0 是一个免费开放源代码的 Web 应用服务器，由 Apache、Sun 公司共同开发而成。是 SUN 公司提供的用于发布信息和将应用程序加载到网站上的服务器，它支持网站创建、配置、管理以及其他 Internet 功能。它具有运行时占用系统资源小，扩展性好，支持负载平衡与邮件服务等特点。Tomcat 与 IIS、Apache 等 Web 服务器一样，具有处理 HTML 页面的功能，另外它还是 Servlet 和 JSP 容器。

Eclipse 是一个开放源代码的、基于 Java 的可扩展开发平台。就其本身而言，它只是一个框架和一组服务，用于通过插件组件构建集成开发环境。利用 Eclipse 进行程序开发，可以根据系统开发的需要，安装相关插件，如 Lomboz，myeclipse 等。

HTML 即超文本标记语言或超文本链接标示语言，是由 HTML 命令组成的描述性文本，HTML 是网络的通用语言，是一种简单、通用的解释性标记语言。HTML 命令可以说明文字、图形、动画、声音、表格、链接等。HTML 的结构包括头部和主体两部分，其中头部描述浏览器所需的信息，主体为页面的具体内容。

JavaScript 是一种由 Netscape 的 LiveScript 发展而来的脚本语言，可以直接把 JavaScript 代码写入 HTML 文档中，浏览器读取时可以进行编译和执行。JavaScript 能使网页增加互动性，使有规律重复的 HTML 部分简化，减少客户端页面的下载时间。通过 JavaScript 还可以将用户验证等操作放到客户端，从而减少服务器的负担。

JSP 类似于 ASP 技术，是一种实现普通静态 HTML 和动态 HTML 混合编码的技术。它是在传统的网页 HTML 文件中插入 Java 程序段（Scriptlet）和 JSP 标记。JSP 能将网页逻辑与网页设计显示和分离，支持可重用的基于组件的设计，使基于 Web 的应用程序开发变得迅速和容易。

Servlet 又称为服务器小程序，是一种独立于操作系统平台和网络传输协议的服务器端的 Java 应用程序，它用来扩展服务器的功能，可以生成动态的 Web 页面。它是位于 Web 服务器内部的服务器端的 Java 应用程序，与传统的从命令行启动的 Java 应用程序不同，Servlet 由 Web 服务器进行加载，该 Web 服务器必须包含支持 Servlet 的 Java 虚拟机，Servlet 在 MVC 模式下充当控制器，能够将模型与视图匹配在一起，并完成用户的请求。

Struts 是 Apache 组织的一个开放源码项目，是一种基于 MVC 模式的 Web 应用，提供了对开发 MVC 系统的底层支持。它能够减少运用 MVC 设计模型开发 Web 应用的时间，它采用的主要技术是 Servlet，JSP 和用户自定义标签等。Struts 框架能大大提高开发效率，同时提高系统的可维护性和可扩充性。

21.3.2　系统运行界面

系统的主要功能实现界面如图 21-5～图 21-8 所示。



图 21-5　水环境预警



图 21-6　非水环境预警



鱼病诊断及预警平台

水环境预警　非水环境预警　疾病预警　返回水产首页

项目	内容
鱼类名称:	牙鲆
请选择:	⊙知道症状 ○知道病名
头部:	□ 眼球白浊　□ 头、口部有白色块状物　□ 头部充血突出
鳃部:	□ 鳃盖内发红　□ 鳃丝发白　□ 鳃褪色　□ 鳃部有寄生虫　□ 鳃盖凹陷　□ 鳃烂　□ 鳃部有白点
体表:	□ 瘦弱　□ 体色变黑　□ 腹部塌陷　□ 鳍边缘呈白浊,变形　□ 体表有白色块状物　□ 腹部膨胀　□ 体表有白色粟状包囊　□ 背鳍有白色块状物　□ 鳍有损坏　□ 体表溃疡　□ 体表有点状出血斑　□ 口部糜烂　□ 胃囊膨大　□ 患处粘液增多　□ 表皮消肿糜烂　□ 皮下大面积出血　□ 体内出现小气泡　□ 脱鳞　□ 肛门扩张发红　□ 肠道脱垂于肛门外　□ 颈部,背鳍下部隆起　□ 鳍发红,出血　□ 体表出现小而隆起的脓肿
内脏:	□ 消化道萎缩　□ 内脏出血　□ 肝脏褪色　□ 肠道白浊　□ 消化道有大量饵料　□ 消化道内充满水样物质或气泡　□ 腹腔有血性腹水　□ 肾脏肿胀　□ 肝脏出血　□ 肾脏或脾脏有白色结节　□ 肝脏老土黄　□ 胃空且肠道浑浊　□ 肠管发炎　□ 消化道内很少有饵料
鱼的行为:	□ 摄食不良　□ 无活力　□ 随水体游动　□ 反应迟钝　□ 受刺激会呈痉挛状　□ 生长缓慢　□ 呼吸困难　□ 竖身游泳　□ 浮于水面打转　□ 鱼本息散乱　□ 水狂游　□ 游动无力　□ 呼吸困难　□ 摇摇晃晃游动　□ 在池中不安，反复窜出水面
镜检:	□ 镜检可发现霉菌体
发病率:	%
死亡率:	%
	查看预警结果

中国农业大学智能化研究所
版权所有 Copyright@2008-2009 All rights reserved.

图 21-7　症状预警

鱼病诊断及预警平台

水环境预警　非水环境预警　疾病预警　返回水产首页

2008年6月24日 星期二 最新警情[低警] 请注意鱼的摄食 2008-05-05

预警结果

项目	内容
鱼病名称:	霉菌病
警级大小:	中警
预警预案:	保持水质良好
鱼病名称:	腹水病
警级大小:	中警
预警预案:	增加换水量,保持水质,洁净的饵料,减少放养密度,土霉素药浴
鱼病名称:	链球菌病

中国农业大学智能化研究所
版权所有 Copyright@2008-2009 All rights reserved.

图 21-8　预警结果界面

21.4 小结

本章从系统需求出发，对系统的结构、功能模块、工作流程和数据库进行了详细设计，配置了系统的软件平台，并采用基于 Struts 的 MVC 开发模式，利用面向对象的思想，以 Windows XP＋Tomcat 5.0＋SqlServer2000 为服务器平台，使用 Eclipse3.2＋MyEclipse 为开发工具开发了集约化养殖牙鲆疾病预警系统，实现了预期功能。

参考文献

1. 殷禄阁，郭金龙. 工厂化养殖牙鲆疾病防治对策[J]. 齐鲁渔业，2007，24(5)：29
2. 林明辉，刘志军，陈总会. 南方工厂化养殖牙鲆防治技术研究[J]. 中国水产，2006(3)：54～55
3. Petra Perner，Silke Jänichen，Horst Perner. Case-based object recognition for airborne fungi recognition [J]. Artificial Intelligence in Medicine，2006，36：137～157
4. Chieh-Yuan Tsai，Chiu CC，Chen JS. A case-based reasoning system for PCB defect prediction[J]. Expert Systems with Applications，2005，28：813～822
5. 杨丽娜，魏永红. 基于 struts 技术的 web 应用设计与实现[J]. 计算机信息与技术，2006，8：35～36
6. 任中方，张华，闫明松. MVC 模式研究的综述[J]. 计算机应用研究，2004，10：5～8，12

第22章

农业病虫害远程诊治与预警平台的设计与实现

22.1 农业病虫害远程诊治与预警平台概述

22.1.1 研究背景与意义

我国是一个病虫害多发的国家,主要农作物病虫害达1400余种,每年因病虫害造成粮食损失约5000万吨,棉花100万吨,蔬菜、果树的病虫害更是严重,同时每年因水产养殖病害问题而造成的直接经济损失就达百亿元之巨。病虫害在导致巨大经济损失的同时,也常造成环境污染与食品污染,对人类食品安全构成威胁。

我国农业病虫害专家比较缺乏,平均每万农业人口中科技人员仅有6.6人,加上农民的科技素质还比较低,绝大多数农民还处在半文盲或文盲状态,对病虫害发生的规律认识不够,因此不能科学地防治病虫害,很难适应病虫害监测、管理及预测防治的要求。为此,如何准确、及时、快捷地监测与诊断病虫害,并提供病虫害的诊治方法,是有效控制病虫危害、保障农业安全的基础性工作。

为了解决我国病虫害发生严重、农业生产分散、病虫害专家缺乏、农民素质低、科技服务与推广水平差等现实问题,设计开发了农业病虫害远程诊治及预警平台。该平台是现代通讯技术、计算机网络和多媒体技术发展的最新成果,养殖户可以通过Web、电话、手机等设备对农业病虫害进行诊断和治疗,同时也可以得到专家的帮助。该平台实现了农业病虫害诊断、防治、预警等知识表示、问题求解与视频会议、呼叫中心、短消息等新技术的有效集成,实现了网络诊断、远程会诊、呼叫中心和移动式诊断决策多种模式的农业病虫害诊断防治体系。

22.1.2 系统实现目标

结合当前农业病虫害发生现状和存在问题以及我国农村信息化程度的调研,系统的总体设计目标如下:

(1) 建立五个诊治平台:Web诊治平台、专家会诊平台、短消息诊治平台、呼叫中心诊治平台、M2M手机在线诊治平台,实现全国范围内广大农户通过多种方式进行病虫害诊治。

(2) 构建农业病虫害预警系统,实现对农业病虫害的早期预警和预报功能。

(3) 构建和开发重大动物疫病预警与疫病咨询信息系统,为畜禽养殖业提供一个有效的动物疫病预警信息平台,实现对重大动物疫情的咨询和知识普及。

农业病虫害远程诊治与预警平台根据诊治种类不同分为五大类:水产品病害诊治及会诊系统、家禽病害诊治及会诊系统、农作物病害诊治系统、果蔬病虫害诊治系统、家畜病害诊治及会诊系统,共完成35个生物品种、646种病的诊治。系统可以诊治的对象如表22-1所示。

表 22-1 农业病虫害远程诊治和预警平台的诊治对象

水产品	七大家鱼(青、草、花、白、鲤、鲫、鳊)
	名贵鱼及观赏鱼(鳊鱼、鲶鱼、鳢鱼、鲮鱼、鳗鱼、金鱼、鲂鱼、鳜鱼、罗非鱼、鳟鱼、雅罗鱼、鲈鱼、白鲳、其他鱼)
	水产动物(牛蛙、蟹、虾、龟、鳖)
家畜	牛、羊、猪
农作物	棉花、水稻、玉米、小麦、大豆
果蔬	梨、桃、苹果、西瓜、黄瓜、番茄
家禽	鸡、鸭、鹅

22.2 系统平台总体设计

22.2.1 系统平台的设计思想

农业病虫害远程诊治和预警平台力求建成包括多种农产品诊断对象,为用户提供多种方式的诊断平台,平台设计遵循以下设计原则:

(1) 架构合理:就是要采用先进合理的软件来设计架构系统,使整个系统安全平稳的运行,并具备未来良好的扩展条件,扩展新功能时,只需添加相应的模块即可实现新的功能。

(2) 方便性:本系统主要服务于广大农户,应提供例如手机、网络、电话等多种方式的使用系统,同时,系统界面应友好,操作方便。

(3) 可扩展性:当增加新的农产品类型时,开发人员可以在此系统基础上进行功能的扩充,而不必进行系统的重构,无需修改系统的部署结构和应用软件,技术架构基本能适应上述变化,保证运转稳定。

(4) 稳定性:系统具有一定的稳定性,有容错处理,当并发访问数量较大时,系统应有自保能力,不应造成系统瘫痪。

22.2.2 系统平台的开发路线

结合技术的可行性与实用性,系统按照需求调研、关键技术研究、总体设计、系统开发和系统应用步骤完成。

在对农户需求分析的基础上,结合文献查阅和专家访谈,对领域知识进行系统分析和收集整理,在充分研究病害诊治模型、病害预警模型、专家会诊算法、领域知识获取模型、知识表现等关键技术的基础上,集成智能代理、远程诊断与呼叫中心技术,开发建立用于水产、家禽、牲畜、果蔬等农业病害的远程诊治平台,实现病虫害的预报、诊断和防治,系统开发技术路线如图 22-1 所示。

22.2.3 系统平台的体系结构

基于上述的系统设计思想,按层次化思路,农业病虫害远程诊治和预警平台的系统体系架构分为五层,由基础平台层、信息资源层、应用支撑层、应用层、访问界面层组成,如图 22-2 所示。



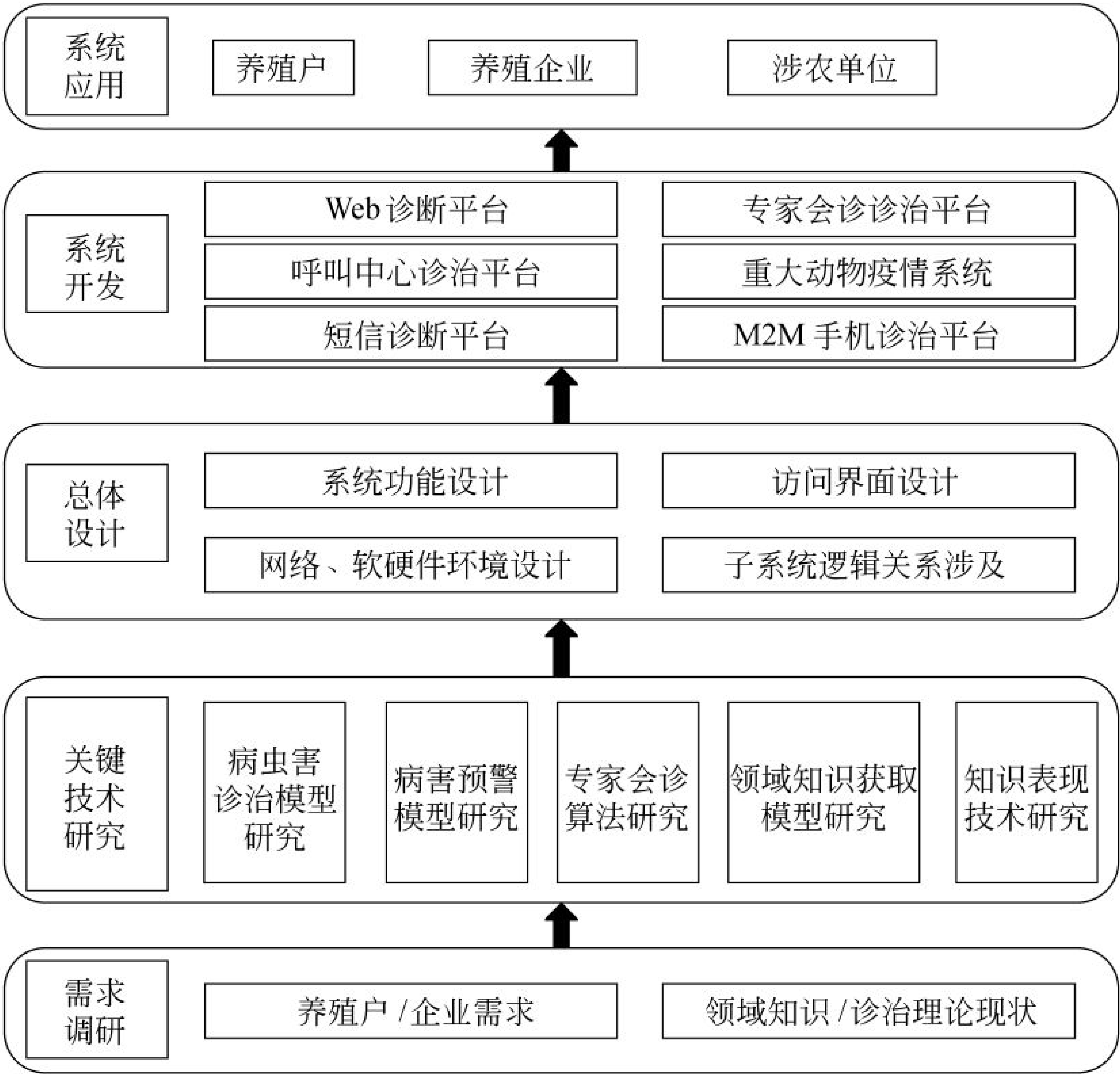


图 22-1　系统开发技术路线

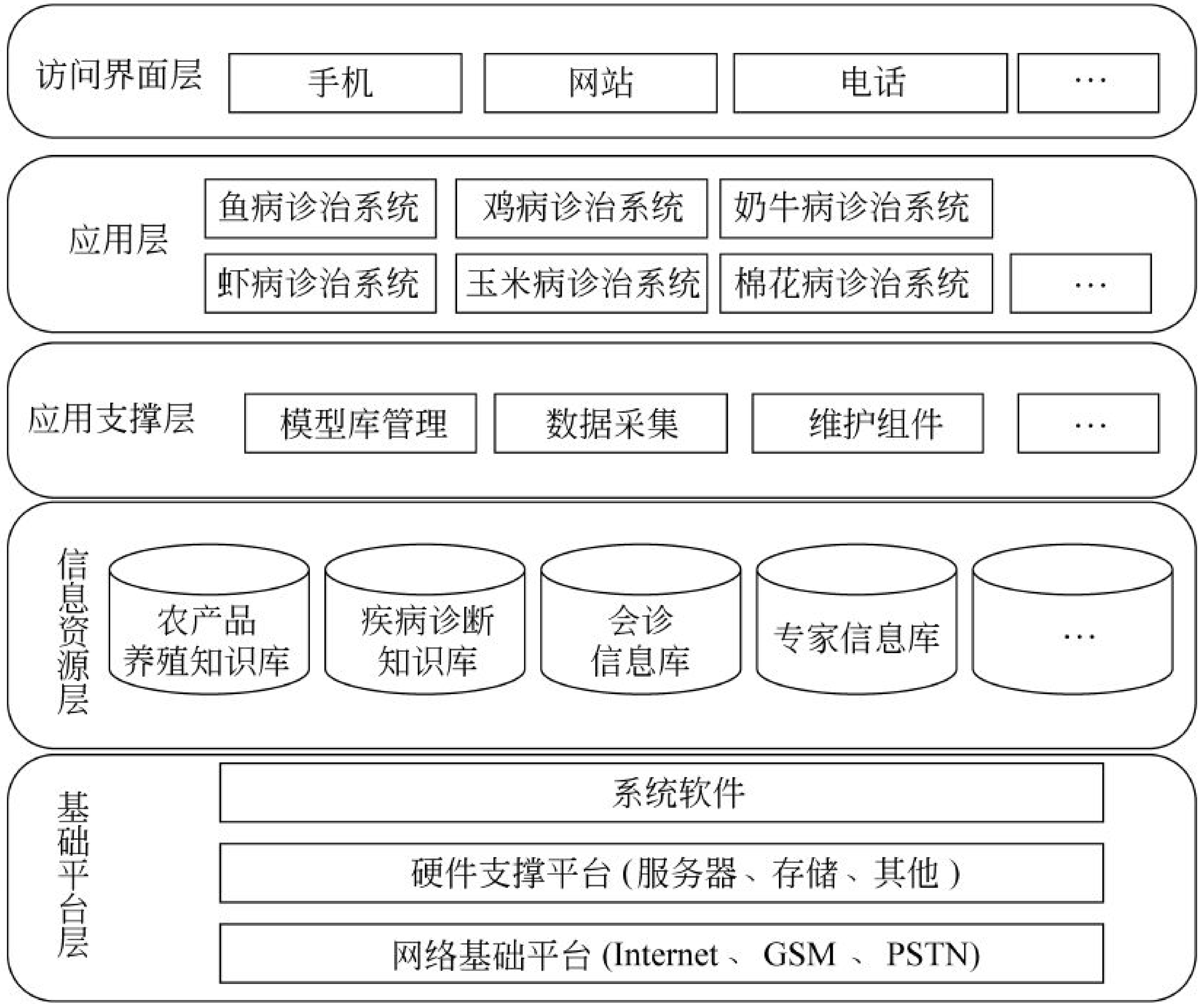


图 22-2　系统体系架构图

访问界面层：该层是直接面向用户的系统界面。用户可以通过多种方式访问系统并与系统交互，访问方式包括手机、网站、电话等。要求界面友好，操作简单。

应用层：提供所有信息应用和疾病诊断的业务逻辑。主要包括分解用户诊断业务请求，通过应用支撑层进行数据处理，并将返回信息组织成所需的格式提供给客户端。

应用支撑层：构建在J2EE应用服务器之上，提供了一个应用基础平台，并提供大量公共服务和业务构件，提供构件的运行、开发和管理环境，最大限度提高开发效率，降低工程实施、维护的成本和风险。

信息资源层：整个系统的信息资源中心，涵盖所有数据。它是信息资源的存储和积累，为农业病虫害诊治应用提供数据支持。

基础平台层：该层分为三部分：系统软件、硬件支撑平台和网络基础平台。其中，系统软件包括中间件、数据库服务器软件等；硬件支撑平台包括服务器、存储等硬件设备；网络基础平台为系统运行所依赖的网络环境。

22.3 农业病虫害诊治系统二次开发平台

22.3.1 开发目的

农业病虫害诊治系统涉及的诊断对象较多，并且诊断对象都通过提供发病症状进行诊治。如果每一种农产品都开发一套诊断系统，不仅工作量大、冗余多且不便于数据管理。为此开发了农业病虫害诊治系统的通用模块，通用模块不仅能对现有的37种农产品进行诊治，更方便后期扩展。只要整理出诊断对象的症状、疾病、防治知识及症状权重列表就可以通过数据库设置来完成其他农产品诊断系统的添加，无需再编写代码，节省了重复开发的时间。

22.3.2 平台功能

使用农业病虫害诊治系统二次开发平台可以直接生成诊治系统，管理员首先整理诊治知识，并将整理的知识倒入数据库，再在门户网站中添加链接即可采用生成的新诊治系统。

（1）整理知识。知识包括：农产品症状知识、农产品疾病知识、防治知识。其中诊治知识主要包括农产品症状部位及症状的具体描述，用以前台显示让用户直观的选择。疾病知识包括疾病的名称及症状权重列表，结合用户选择的症状判断农产品所患疾病。防治知识主要用于诊断后提供给用户一些专业知识，如疾病定义、疾病治疗、疾病预防等。

（2）将整理的知识导入数据库，并在配置表里配置新加的农产品信息。其中配置主要包括：农产品名称、农产品系统的banner图片路径、农产品系统配色等。再调用程序生成模拟案例表。

（3）添加访问链接至总诊治系统门户网站首页。

22.3.3 开发架构和模式

平台开发采用MVC模式，采用Struts的体系结构（如图22-3所示）。Struts的体系结构实现了MVC设计模式，利用Struts可以非常快速构建一个结构清晰、可扩展性较强的系统。

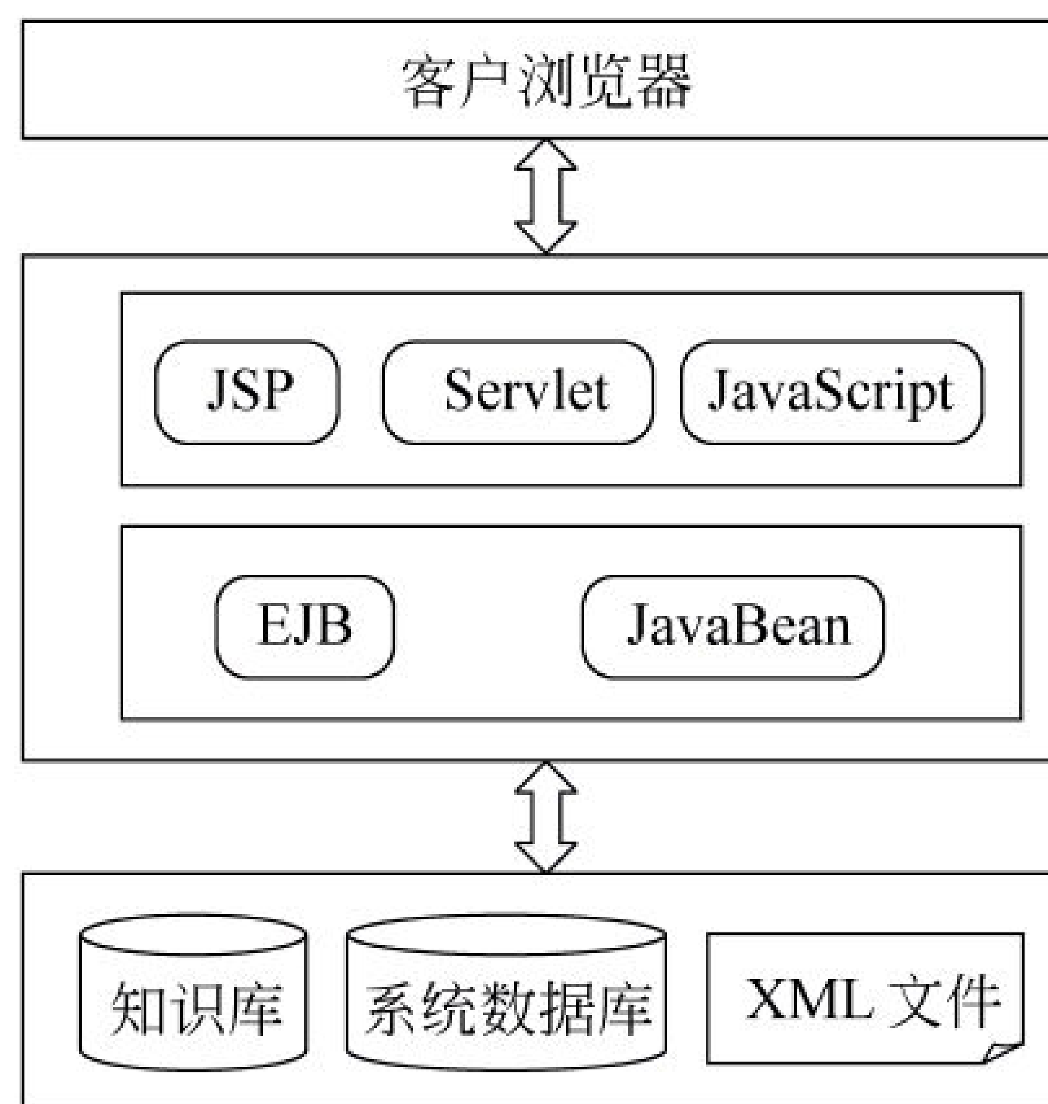


图22-3 系统开发架构

（1）客户层：系统用户和普通用户通过Web浏览器访问系统和社区系统。系统用户通过Web浏览器定制自己的社区子系统，普通用户通过Web浏览直接浏览生



成的社区子系统。

(2) 表示层：表示层主要是指封装在Web容器中的JSP页面和Servlet。

(3) 业务层：本系统中，执行业务逻辑的包括普通业务Bean和EJB，是整个系统的支撑平台，表示层根据客户端的需求，通过创建构件实例，调用方法，修改方法，根据需要使用这些构件来完成指定的业务。这一层是整个项目的核心部分。因为要允许二次开发者定制这些构件，所以构件暴露给开发者一些接口，通过这些接口完成对构件的定制或扩充。

(4) 数据层：负责系统结构化数据和非结构化数据的存储、检索、优化。对于结构化的数据，通过硬件系统配置，能够实现自动备份、故障诊断、自动恢复，为数据的存储性能和安全提供全方位的保证。

22.3.4　系统实现

农业病虫害远程诊治和预警系统采用MVC开发模式，该模式能够有效存储和展示数据对象，降低模块之间的耦合度，将传统的输入、处理和输出模型转化为图形显示的用户交互模型。MVC的设计模型由三部分组成：控制器、视图、模型，该模式能够明确各部分的责任，减少代码的重复性。系统开发界面如图22-4～图22-7所示。



图22-4　农业病虫害诊治及预警系统

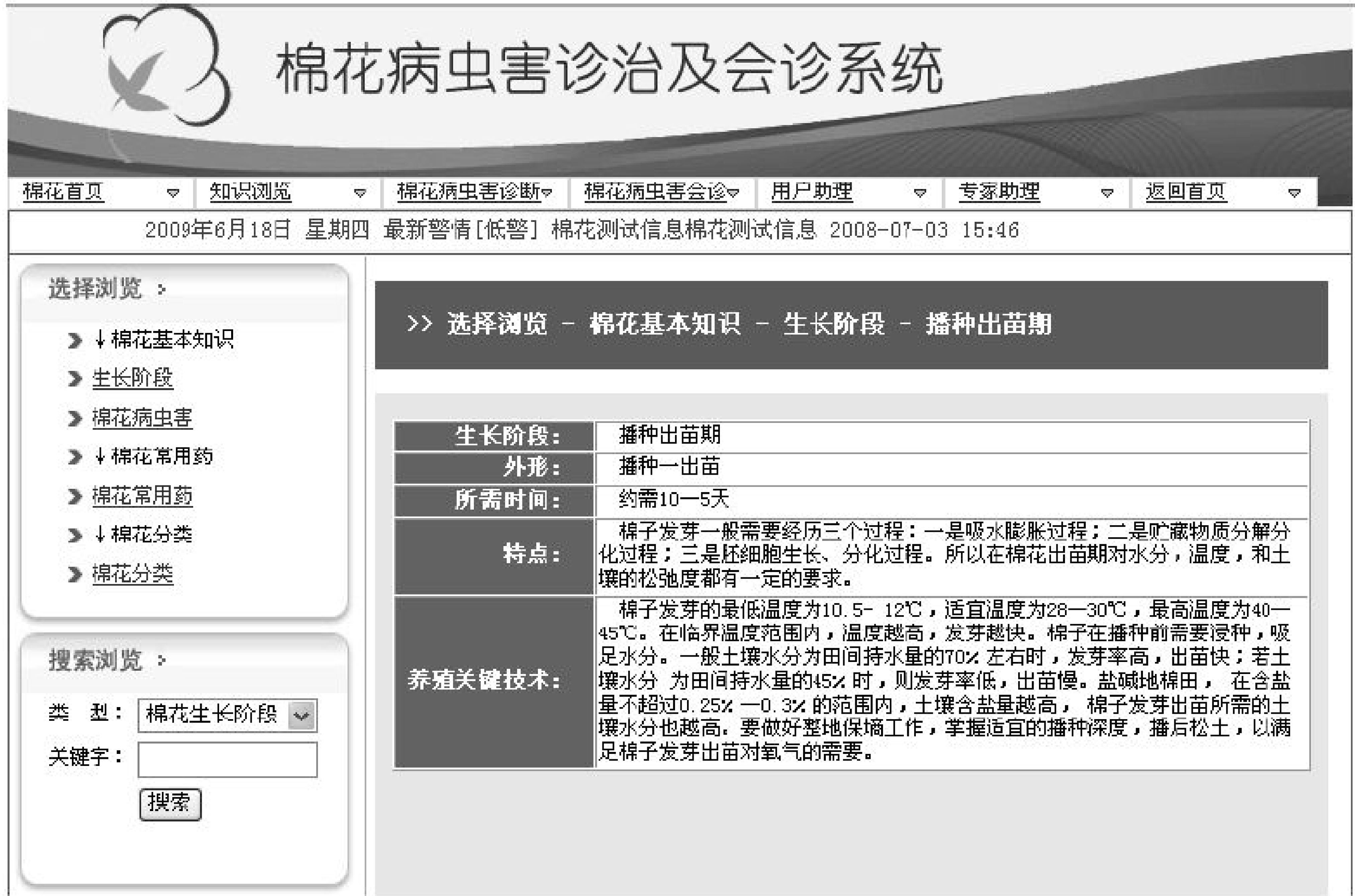


图22-5　棉花病虫害诊治及会诊系统



图 22-6　牛病诊治及会诊系统



图 22-7　西瓜缺素症诊治系统

22.4　农业病虫害诊治系统的通用维护平台

22.4.1　开发目的

由于农业病虫害诊治系统需要大量的知识表来支持，如农产品养殖知识表、病害诊断知识表等。为了能统一管理数据，方便维护人员对数据的维护，针对农业病虫害诊治系统开发了数据库通用维护平台。只需要系统管理员定义好数据库的结构及属性，维护人员就可以轻松地维护数据。同时可以给维护人员分配角色，控制维护人员能够维护指定的数据。在维护数据



的同时保证数据的安全性。

22.4.2　平台功能

通用维护平台用户分为两种：系统管理员和数据维护人员。

(1) 系统管理员功能及实现：模块管理、数据表管理、数据关联关系管理

模块管理实现对数据表归类的目的，便于维护人员在数据维护时能更快更方便地找到要维护的数据库，这通过数据库模块表和数据表之间的表关系实现。数据表管理分为数据表信息管理及数据表下属列管理。数据表信息管理用于维护数据表信息以便前台显示，数据库表功能属性设置，如是否可删除等操作。下属列的管理基本和表操作类似，可分列信息维护及列功能属性操作，通过对数据表及下属列表操作实现。数据关联关系管理是对下属列数据来源的设置管理。数据列数据来源分别为实际值和关联表(见图22-8)。

图22-8　系统管理员功能页面

(2) 维护人员功能及实现：数据表数据的增加、删除、修改

维护人员可以对数据进行添加，点击添加按钮后，程序将查询数据库信息，获取该数据表的信息及功能属性，并返回页面以表单形式显示出来。当用户点击提交后，程序再次查询数据库获取数据表的验证信息，对数据进行验证，验证合法后添加至数据库。

修改操作和添加功能相似，但有别于添加功能的是以表单形式显示的同时显示历史数据，以便用户对比修改。

删除功能通过程序调用数据表的信息及功能属性。首先通过功能属性判断数据表是否能删除，再进行删除或提示操作(见图22-9)。



水产品名称	别名	操作	操作	操作
草鱼	混子、鲩鱼、草青	添加	删除	修改
青鱼	黑鲩、青鲩、青根鱼	添加	删除	修改
鲢鱼	白鲢、鲢子	添加	删除	修改
鳙鱼	花鲢、黄鲢、胖头	添加	删除	修改
鲤鱼	龙鱼、花鱼、红鱼、鲤 ...	添加	删除	修改
团头鲂	武昌鱼、团头鳊	添加	删除	修改
鲫鱼	鲋鱼、刀翅鱼、喜头	添加	删除	修改
罗非鱼		添加	删除	修改
翘嘴红鲌	翘嘴鲌丝、大白鱼	添加	删除	修改
鳜鱼	季花鱼、鳌花鱼、旁鳜 ...	添加	删除	修改
鲴属鱼类	银鲴、细鳞斜颌鲴、黄 ...	添加	删除	修改
牙鲆	偏口鱼、牙鳎、花瓶、 ...	添加	删除	修改
半滑舌鳎	龙力、舌头、牛舌、鳎 ...	添加	删除	修改

图 22-9　数据维护人员功能页面

22.5　基于机器到机器通信的农业病虫害诊治系统实现

22.5.1　机器到机器通信技术介绍

机器到机器通信(machine to machine，M2M)，通过在机器内部嵌入通信模块，采集机器的运行状态后用移动无线网络转递给后台应用程序进行处理，从而实现对机器的远程监控与管理，满足客户对生产过程监控、指挥调度、远程数据采集和测量、远程诊断等方面的信息化需求。系统具有以下特点：

- 无线方式传输数据、避免布线、节约成本；
- 用户可现场实时采集数据，数据时效性强；
- 无需人工干预，实现数据自动上传，提高信息处理效率。

通过与中国移动通信集团重庆有限公司合作，设计开发了基于 M2M 的农业病虫害远程诊治和预警平台，将 M2M 技术首次应用到农业病虫害的诊治过程中，系统说明如图 22-10 所示。

22.5.2　系统功能

当鱼类生活的水质被污染、溶解氧过低，鱼类生病或药物中毒，健康状况下降，出现鱼体不适或死亡时，鱼会侧翻，腹部颜色较浅部分会浮在水面上，与水池较深的背景颜色有较大的区别，容易区分。并且在养殖场中统一放养的鱼具有品种相同，大小相近等特点，这就可以利用统计方法，得出分割病鱼与池塘水背景色的阈值，同时也可以得到某一段时期池塘内鱼的大小及腹部白色区域的大小。因此，只要将得到的原始图像二值化，将鱼体的白色腹部与背景分割开，然后统计白色区域部分的数目和大小，当最小白色区域大于等于该段时间内统计值的下



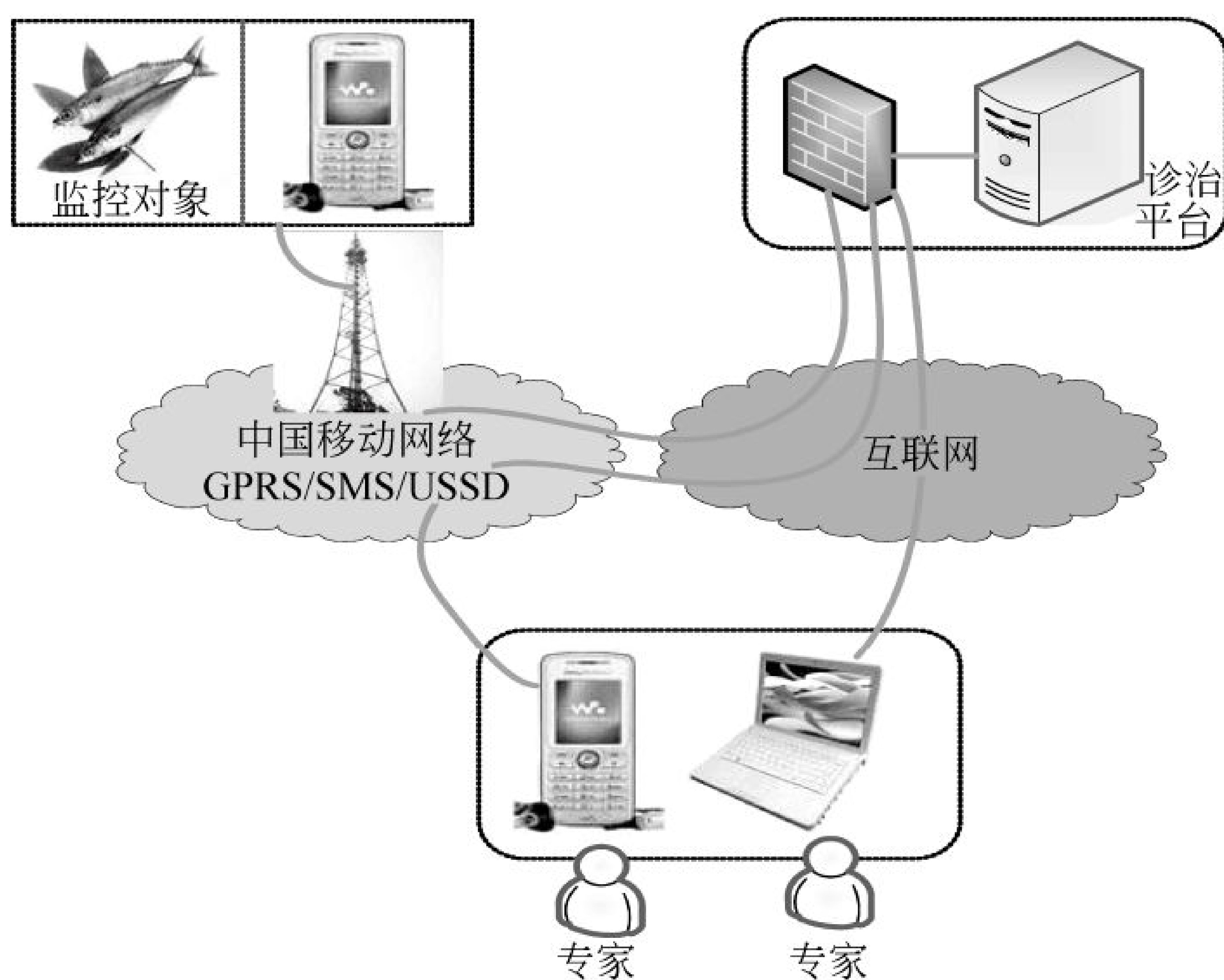


图 22-10 M2M 系统架构

限，且小于统计值的上限时，即认为有鱼出现不适反应或已经死亡，需要早期预报，采取措施救治。

该图像处理系统由用户采用手机拍摄现场，通过无线网将图片上传到系统服务器，系统服务器对图片进行分析，得出诊断结果，可以采用短信方式将诊断结果传递给用户。从而实现早期预防和报警处理。

本系统的优点是把计算机的图像处理方法应用于水产养殖的鱼病自动诊断中，可以对鱼病进行早期诊断和预报，将信息及时上传给上位机，并以短信形式通过移动通信网发送给管理人员，以便及时处理。

22.6 基于短信的农业病虫害诊治平台

基于 Web 的农业病虫害诊断系统在一定程度上解决了病害频繁发生而领域专家缺乏的矛盾，减少了病害带来的经济损失。但是由于我国农村信息基础设施薄弱，很多地方不具备计算机和互联网服务条件，因此基于 Web 的诊断系统的使用受到了一定限制。

短消息业务(short message service，SMS)具有覆盖范围广、费用低廉、稳定性高等优点，适合应用于频繁、远距离传送小流量数据的领域。中国农村目前有 4000 多万手机用户，并且每年以 30%以上的速度增长。因此，建立基于短信平台的农业病虫害远程诊治和预警平台对探索我国新农村信息服务形式，提高信息服务质量，解决农村地区分布广且相对偏远以及农业专家相对缺乏并集中于城市的矛盾，具有积极的现实意义。

22.6.1 系统结构设计

系统是主要由 GSM/GPRS 模块、短信诊断平台(诊断服务器)和用户端(PDA、CDMA 手机以及小灵通)组成，如图 22-11 所示。GSM 模块通过数据线与计算机的 USB 接口连接，SIM 卡安装在 GSM 模块中，由计算机软件进行短信的收发和管理。在病虫害诊断过程中，用户需要在短信平台的提示下不断地和诊断短信平台进行信息交流，通过短信平台进行疾病诊断的

过程就是用户与专家系统不断交互的过程。短信平台的操作主要包括：

（1）设置短信平台端口、波特率、授权号码以及输入使用的 SIM 卡卡号。

（2）测试短信平台功能：对短信平台接收短信和发送短信的功能进行测试。

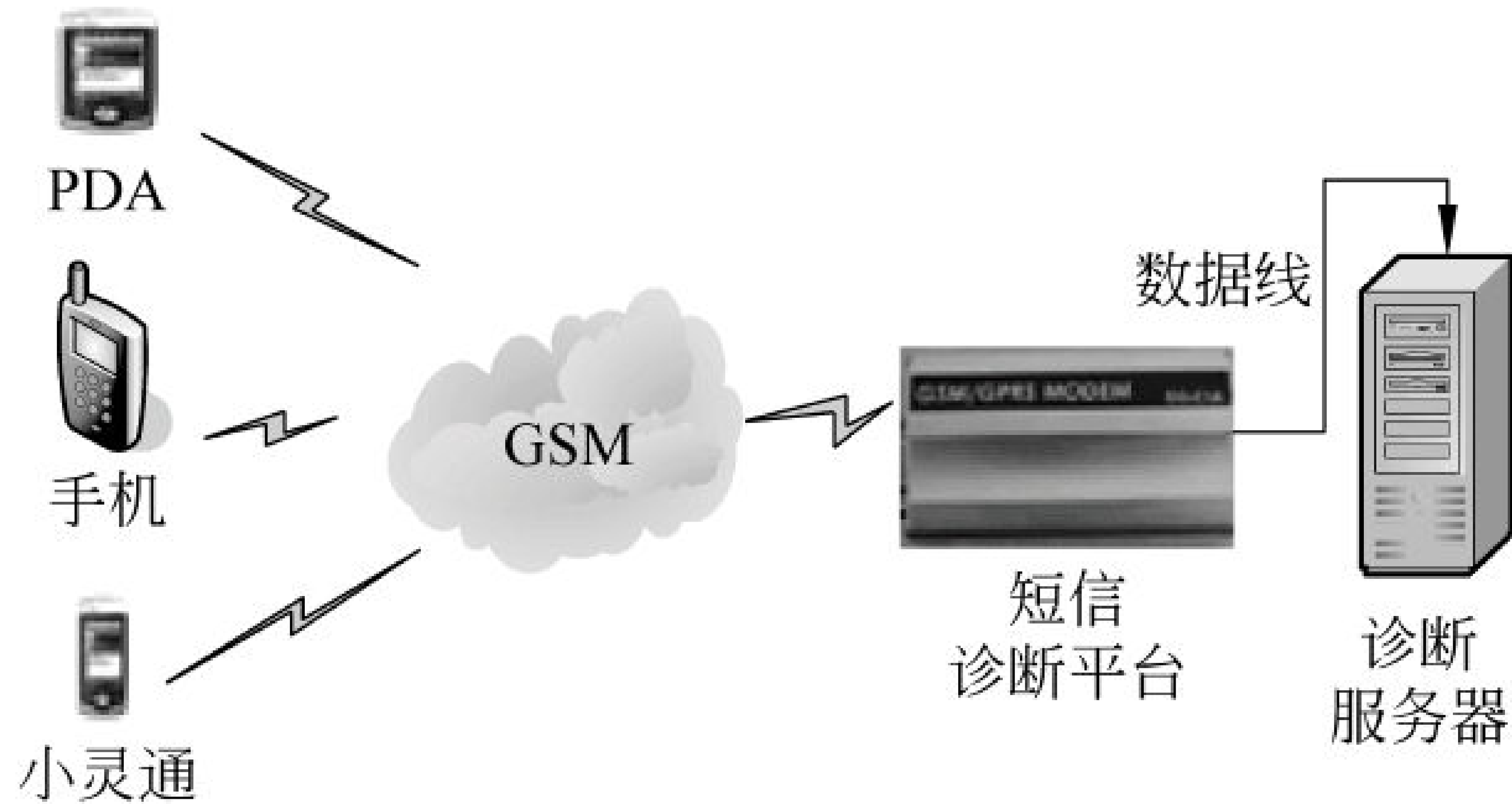


图 22-11　基于短信平台的农业病虫害的远程诊治系统

22.6.2　体系结构设计

系统采用“应用层-通信层-逻辑层-数据层”四层分布式软件体系结构，如图 22-12 所示。

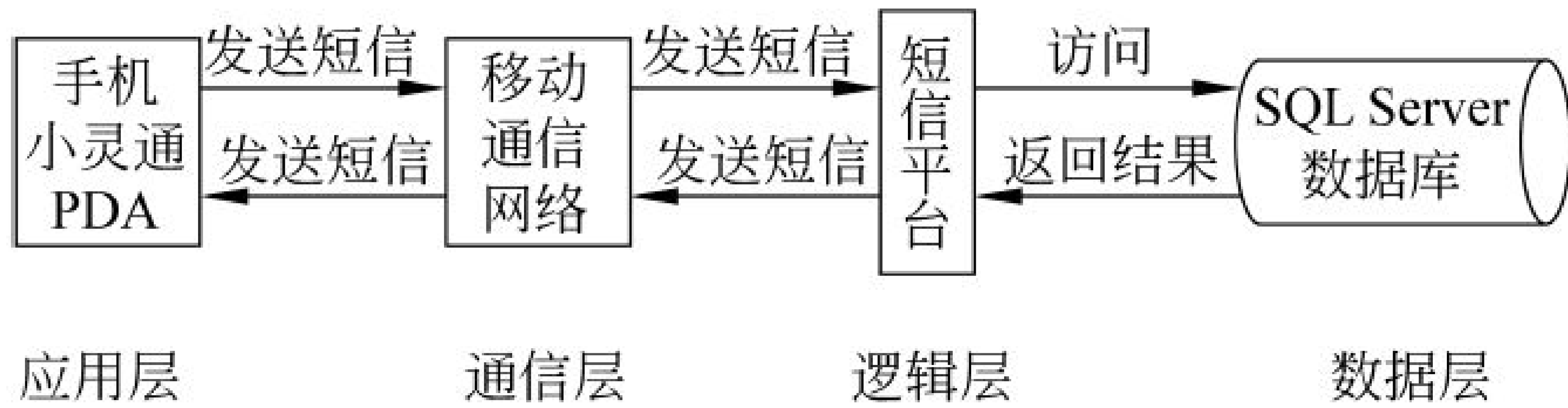


图 22-12　系统软件体系结构图

（1）应用层：应用层主要是指用户与系统交互的平台，在本系统中主要指移动通信设备，如手机、PDA、小灵通等，是用户向系统请求诊断、输入病虫害信息以及接收诊断结果的工具。

（2）通信层：通信层是信息交换的媒介——移动通信网络，提供移动服务，互联应用，无线网络建设，GSM 短信息应用等，在本系统中负责传递客户层和诊断短信平台之间的信息。

（3）逻辑层：逻辑层一般也称业务处理层，处于通信层和数据层之间，是一种典型的中间件技术，负责核心业务的处理。逻辑层主要对客户层发送过来的病虫害诊断信息的内容进行判断，完成事务的逻辑处理，并对用户的输入进行诊断推理，将诊断结果和防治措施结果返回给用户。

（4）数据层：数据层也称数据服务器层，一般由大型的数据库服务器系统承担，负责系统结构化数据的存储和管理，主要包括诊断对象的基本知识、病虫害的症状信息、诊断案例等。

22.6.3　系统实现

本系统主要实现手机用户与短信平台的自动交互功能。用户首先输入“诊断”的汉语拼音首字母 ZD，以此作为用户开始诊断的标志，用户发送 ZD 到农业病虫害诊断短信平台，系统则对用户的诊断请求进行处理并到定制的短信平台请求诊断，然后根据短信平台的提示，输入对



应疾病的症状；短信诊断平台记录用户输入并采用基于案例的算法进行诊断推理，然后将诊断结果和防治措施反馈给用户。

与网络诊断不同，短信容量较小，不能接收和输出较多的内容，诊断信息的提交也不能一次完成，这就要综合一条手机短信的容量和诊断过程中提交的疾病信息来设计短信诊断平台的诊断过程和系统编码。系统的诊断流程如图22-13所示。

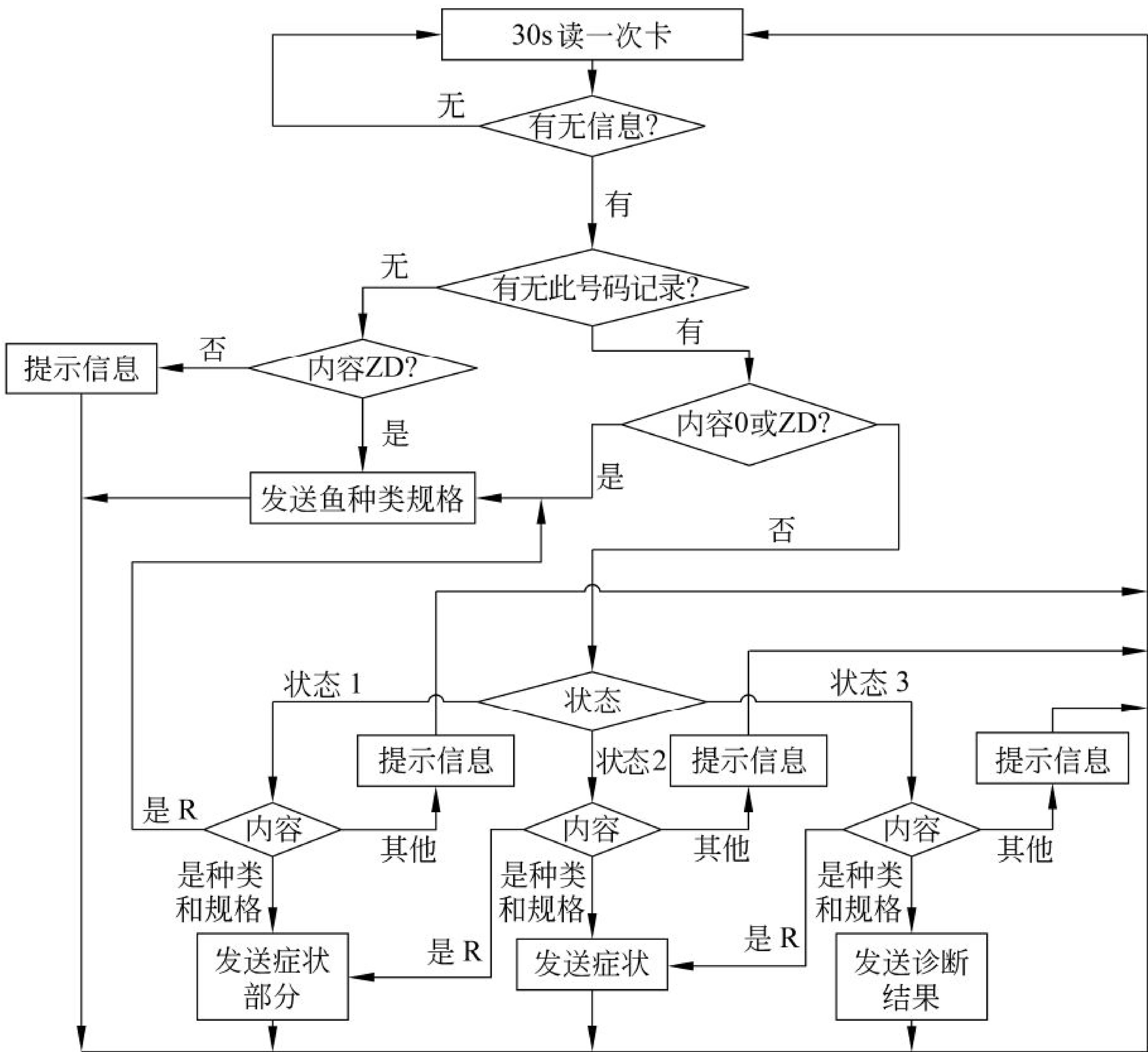


图22-13 基于短信平台的病虫害诊断流程

22.7 小结

为了满足不同用户的使用需求，农业病虫害远程诊治和预警平台一体化了基于Web、短信平台、呼叫中心、专家会诊、M2M手机在线诊治五种不同的诊断模式，大大降低了系统应用的条件，提高了系统的实用性和覆盖范围。

上述一体化平台基于病虫害诊断理论、预警理论和群决策理论，在农业病虫害诊断知识获取、表示、诊断推理以及预警方面作了积极探索，并集病虫害预警、诊治、专家会诊多种功能为一体，为用户提供多方位的服务。

该平台集成网络、电话、手机、短信等多种交流方式，可以进行全天候农业病虫害诊断服务咨询，对探索我国新农村的信息服务形式，提高信息服务质量，解决农业养殖和种植区域分布广且相对偏远与农业专家相对缺乏并集中于城市的矛盾，具有积极的现实意义。

参考文献

1. 祝伟. 基于呼叫中心的鱼病智能诊断系统[D]. 中国农业大学,2006
2. 温继文. 基于知识推理的鱼病诊断系统研究[D]. 中国农业大学,2003
3. 杨萍. 鱼病远程会诊决策支持系统[D]. 中国农业大学,2006
4. 胡利平. 基于智能计算的移动式专家系统 [D]. 中国农业大学,2005
5. 周云. 基于案例推理的鱼病诊断专家系统 [D]. 中国农业大学,2004

示范篇

第23章　农业病虫害远程诊断技术研究与示范

第23章

农业病虫害远程诊断技术研究与示范

“农业病虫害远程诊断技术研究与示范”项目属于国家高技术研究发展计划(“863”计划)现代农业信息技术专题项目,其研究成果具有可视化、智能化、构件化的特点,所包含的农业病虫害知识领域广泛、信息量大、操作简单,同时具有可延伸性、可维护性和安全性强、在线直观等特点。该系统的推广与示范为努力实现农业由数量型向质量型转变,提升农业病虫害防治的技术与管理水平,降低农业生产成本,提高农民收入,促进我国农业信息化发展打下了良好的基础。

23.1 项目的研究示范内容

23.1.1 研究示范的总体目标和水平

在重大技术层面,实现农业病虫害诊断、防治等综合知识表示与问题求解模型及不确定知识推理算法与智能代理、视频会议系统、呼叫中心、GIS等新技术的结合,并取得重大突破。项目的整体技术水平达到国际先进水平,部分关键技术达到国际领先水平,填补了该领域国际和国内研究的空白。

在重大产品层面,申请有独立知识产权专利软件6～8个,建立适合我国特殊国情的农业病虫害远程诊断体系,构筑我国农业病虫害远程诊断平台3个。

在基地建设方面,实现网上决策系统、远程会诊、呼叫中心和移动式诊断决策系统4种模式的农业病虫害预报、诊断、防治,建立一种24小时全天候服务咨询模式,全面提升我国农业科技服务、推广和应用水平,并建立农业病虫害远程诊断试验示范基地3个;探索我国农业病虫害远程诊断服务与咨询的新模式。项目拟在北京、天津、上海、山东、浙江等推广示范,年用户数达到3万以上,诊断准确率达到90%以上,指导应用面积累计500万亩以上。

23.1.2 项目研究示范的主要内容

1. 农业病虫害基础数据库的构建与完善

在项目各参加单位原有的工作基础上,进一步完善和补充已建立的各种主要农作物病虫害多媒体特征数据库,包括病虫害的症状描述、发病条件、为害状况、防治方法等;信息的表现形式包括图、文、声、像等多媒体形式;建立水产、蔬菜、果树病虫害标准图像库;对上述内容提供方便的多路径检索及复合检索功能。

主要技术指标与水平:农作物(包括水产、蔬菜、果树等)主要病虫害特征库包括300多种主要作物病虫害的基本信息,收集各种病虫害特征图片15000张;标准图像库包括300多种主要作物病虫害的标准图像1500多幅;全部数据库内容数据量在20G左右。

2. 基于B/S模型的农业病虫害诊断平台研究

多Agent病虫害诊断决策系统研究：根据病虫害症状与病因的产生机理，建立反应式Agent的一般病虫害诊断问题的求解模型和不确定知识推理模型；根据农业病虫害发生的一般规律，建立病虫害的预报、诊断与防治概念模型与表示模型，构造完善的知识库；根据知识库完善程度，建立前期学习、后期学习及专家学习机，提高病虫害诊断Agent与学习环境的交互能力；根据病虫害预报、诊断与防治的一般过程，建立便捷、快速、友好的人机界面；根据呼叫与咨询系统的需要，建立人与人交互界面。

水产病害诊断决策系统研究：建立9种常见鱼类、蟹、甲鱼等特种疾病症状、病因、病名与防治方法关系模型，提出鱼病预报、诊断、防治的知识表示方法；运用统计概率和主观概率相结合的方法，提出适合水产养殖业疾病诊断的不确定知识推理算法；根据鱼病预报与诊断问题的求解模型和不确定知识传递算法，建立现场诊断、目检、水质检测、镜检的鱼病预报与诊断推理机。

蔬菜病虫害诊断决策系统研究：建立常见蔬菜病虫害的知识表示模型，该模型重点体现病虫草为害症状、发生特性、防治技术特性；建立蔬菜病虫草害与各种杀虫剂、杀菌剂、除草剂的关系模型；从蔬菜病虫草害识别、发生特性诊断、防治决策方法、治理技术四方面建立蔬菜病虫害不确定性知识求解模型；根据知识表示模型、关系模型和问题求解模型，研制开发主要蔬菜病虫害诊断决策系统。

果树病虫害诊断决策系统研究：根据果树病虫害发生的一般规律，建立多种常见病虫害、症状与防治方法关系模型，提出果树病虫害预报、诊断、防治的知识表示方法；运用统计概率和主观概率相结合的方法，提出适合果树业病虫害诊断的不确定知识推理算法；根据知识库完善程度，建立前期学习、后期学习及专家学习机，提高病虫害诊断与学习环境的交互能力。

主要技术指标与水平：鱼病诊断决策系统将对9种鱼类，150余种常见鱼病进行诊断，鱼病知识库含500余条规则和1000余张特征图片；蔬菜病虫害诊断决策系统将对黄瓜、番茄、辣(甜)椒、茄子常见的60余种病虫害进行诊断，知识库将含300余条规则和800余张特征图片；果树病虫害诊断系统含500条规则和1000余张特征图片，对苹果、梨、桃、葡萄等水果的80多种常见病害及虫害进行诊断与防治。诊断系统诊断与识别的准确率在90%以上。

3. 农业病虫害远程会诊平台研究

实现用户与专家的预约，主持专家对诊断成员的邀请，预约信息的登记及管理；实现对病虫害主要症状图文信息的收集、整理并建立相应的数据库管理系统；会诊结论的收集、整理与诊断结论信息的用户输出。

开发异地呼叫组件，实现农户对专家的呼叫；开发语音交互组件、视频交互组件、文本交互组件，实现异地专家和农户的“面对面”实时交流；开发白板模块，实现专家与异地用户计算机资源共享，实现共享画面读、写、存储、打印等多种方式的实时讨论。

主要技术指标与水平：会诊系统平台1个，实现多点语音、视频、文本、图像的异地实时交互；会诊系统稳定可靠。

4. 农业病虫害远程呼叫与咨询系统研究

实现用户信息档案资料记录和各种数据的统计分析等，供坐席人员在线使用并给专家提



供参考;根据一定的分配算法,将用户呼叫信息(电话访问、IP访问等)接入信息中心服务系统,并合理地分配给后台的坐席处理人员;在用户接入咨询中心时,实现自动语音导航服务。

达到的目标:通过对普通电话、IP电话、Web主页、E-mail等多种媒体的统一受理,最大程度地提高传统的基于中心数据库的专家系统的实时交互性。通过实时交互和非实时交互两种方式最大限度地利用传统专家系统满足农户需求。将专家系统与农户资料信息库高度融合,个性化地根据农户历史资料信息和专家知识系统满足农户需要。及时将农户的一手资料通过专家整理扩充到专家知识库中,最大限度、最快速度地丰富专家知识库中的知识信息。

主要技术指标与水平:在多种媒体多个用户同时服务过程中系统稳定,无数据串音和相互干扰;农户资料信息和咨询系统结合,保证坐席服务过程能及时快速获得用户基本信息和历史咨询情况,对农户实现个性化服务。

5. 移动式农业病虫害诊断决策系统研究

进行农业病虫害诊断决策系统的知识库和推理机研究;实现服务器、工作站数据库、知识库及推理系统向掌上电脑系统的移植;实现田间数据采集、传感系统与掌上电脑的有效连接。

主要技术指标与水平:利用移动式诊断决策系统实现田间病虫害问题的现场诊断;研制常见鱼类、蔬菜和果树病虫害3大类便携式诊断决策系统系列。

6. 基于GIS的农业病虫害广域分析

实现地理空间数据库与农业病虫害诊断知识库、标准图像库和诊断模型、农业害虫的数字化特征库、病虫害知识库等动态集成。包含空间信息(经纬度、行政区划位置等描述性空间信息表示)的病虫害分布信息的获取,对病虫害信息进行广域分析。

预期达到的目标:地理空间数据库与农业病虫害的诊断知识库、标准图像库和诊断模型、农业害虫的数字化特征库、病虫害知识库等动态集成,利用WEB-GIS的方法实现显示,配合诊断决策系统实现田间病虫害问题的诊断;通过使用WEB-GIS,实现对病虫害信息的广域分析。

主要技术指标与水平:地理空间数据库与农业病虫害的诊断知识库、标准图像库和诊断模型、农业害虫的数字化特征库、病虫害知识库等动态集成性能稳定,调用与查询快速、匹配准确;通过病虫害信息的广域分析,为广域病虫害的预报与预防提供参考和依据。

7. 害虫远程自动识别与诊断平台

运用昆虫图像处理及分析系统(BugVisux)对害虫图像进行分析、处理,获取图像的数字化特征值,建立主要农业(如蔬菜)害虫的数字化特征库。运用当前主流的网络技术和通信技术,开发昆虫远程自动诊断与识别软件,该软件包括客户端、服务器端和数据库端三层结构,可通过USB接口的摄像头自动获取图像视频,经过远程传送到达服务器端,由服务器端进行图像分析与处理,最终达到能够对昆虫进行识别诊断的目的。

主要技术指标与水平:构建主要蔬菜、果树等农业害虫的数字化特征库,作为本研究的基础;建立农业害虫自动诊断模型,开发出农业害虫远程自动诊断平台;对开发出的农业害虫远程自动诊断平台进行软件登记。

8. 温室番茄生长建模与智能管理的研究

计算番茄在各个生长阶段形态结构的形态发生模型;模拟温室番茄结构与功能有机结合的机理性生长生理生态模型;提出温室番茄生长全过程的管理决策模型,利用传统或现代的优化算法,实现种植的低成本、高效益;利用虚拟现实建模语言 VRML 将生长模型的计算结果可视化,以更有效地观察与总结番茄的生长发育规律。

达到的目标、主要技术指标与水平:开发一个基于模型的温室番茄生长智能管理软件;建立的番茄模型和软件易于推广应用到其他品种的蔬菜,如黄瓜、甜椒等,并实现与平台的集成;通过实现节水、节肥和节能的多目标优化,最终达到使资源的利用率提高 20%,农业生产成本降低 10%,效益提高 20%的效果。

23.2 项目的完成情况

从整体上看,项目研究按照预定计划正常进行,实现了农业病虫害诊断、防治等综合知识表示与问题求解模型及不确定知识推理算法与视频会议系统、呼叫中心、GIS 等新技术的有效集成;建立了适合我国特殊国情的农业病虫害远程诊断体系,构筑了我国农业病虫害远程诊断平台 3 个;实现了网上决策系统、远程会诊、呼叫中心和移动式诊断决策系统 4 种模式的农业病虫害预报、诊断、防治,建立了一种 24 小时全天候的服务咨询模式。在项目研究过程中,申请有独立软件著作权 18 项,发表了三大检索(SCI、EI、ISTP)收录的论文 15 篇/次,建立了具有“863”品牌与特色的农业病虫害远程诊断研究实验室 3 个。

在北京、天津、山东、福建、重庆与新疆 6 省市进行项目示范,年用户数达到 3 万以上,诊断准确率达到 90%以上,指导与辐射应用面积累计 500 万亩以上。

23.2.1 农业病虫害基础数据库

我们进一步完善和补充了已建立的各种主要农作物病虫害多媒体特征数据库;开发农作物病虫害基础数据库系统,实现了对不同形式(包括文本、图像、声音等)的各种信息的有效组合,建立了多路径组合查询,数据库信息基本能够满足用户查询的需要。

具体工作有以下方面:

(1) 农业病虫害基础数据库扩充与完善。在项目各参加单位原有的工作基础上,收集了农业(包括水产、蔬菜、果树等)包括 300 多种主要作物病虫害的基本信息,主要描述病虫害的形态特征、危害状等,文本信息主要描述病虫害的中文名称、英文名称、学名、寄主植物、形态或症状、发生条件、为害状况和防治方法等;主要病虫害特征库收集各种病虫害特征图片 15 000 张。

(2) 农作物病虫害基础数据库管理系统。开发了基于 JSP+MYSQL 的农作物病虫害基础数据库系统(包括主要农作物病虫害多媒体网络信息系统、主要农作物病虫害标准图像库),实现了主要农作物图文信息的远程检索和远程管理。

(3) Smart 元搜索引擎(Smart Meta Searcher,SMS)系统。采用用户行为分析技术,对检索的关键词进行提取加工处理,选择最优的搜索引擎组,按各搜索引擎的特点将其转换为更精确的检索条件,大大提高了检索的准确率。同时利用高效的融合算法对返回结果进行分析处

理，用星级排行因子标注每条检索结果以描述检索返回的文档与用户所要查找的目标文档的差距，提高了检索的查全率、查准率并具有较好的排序。关键工作主要集中在搜索引擎选择、结果融合算法、用户行为分析与星级排行评价因子等方面。

23.2.2　基于 B/S 模型的农业病虫害诊断平台研究

我们建立了农业病虫害知识表示与推理机制，实现了知识库的优化；设计并实现了农业病虫害诊断平台框架；在平台基础上，结合示范区特点，进一步开发了水产（鱼病、特种鱼、甲鱼）、棉花、家畜（猪、牛、羊）、蔬菜、果树等疾病诊断专家系统，基本上达到用户的使用需求；准备进一步开发家禽疾病诊断专家系统。目前系统准确率在 80%左右，有待于进一步测试与完善。相关内容详述如下。

（1）农业病虫害诊断领域模型与形式化体系。以鱼病诊断为例，结合鱼病诊断学和鱼病专家的知识与经验，建立了包括诊断主体、诊断客体、诊断要素、诊断结果、诊断思维、诊断方法等因素的鱼病诊断的领域知识模型；从领域知识模型出发，将鱼病知识划分为事实性知识、经验性知识与决策性知识 3 类；基于以上分类，初步建立了包括 2 个前提、7 个定义和 3 个推论的鱼病诊断知识概念化体系。该模型的建立不仅可以为农业病虫害远程诊断平台的开发、实现平台的可重用性提供理论基础，而且还可以为选择鱼病诊断知识表示模式提供有效的理论依据，以便更科学、系统地形成鱼病诊断专家系统知识开发及其智能化的科学过程，把这方面工作由分散、随意性的状态，提升到一个更理性、程序性和系统性的阶段。

（2）农业病虫害诊断知识表示的研究。在领域概念模型指导下，采用了基于产生式规则、基于案例和基于二叉树结构的多种知识表示方式的混合模型（见图 23-1），有效解决了仅采用传统产生式规则带来的规则数量过多所导致的推理速度慢、效率不高等问题。

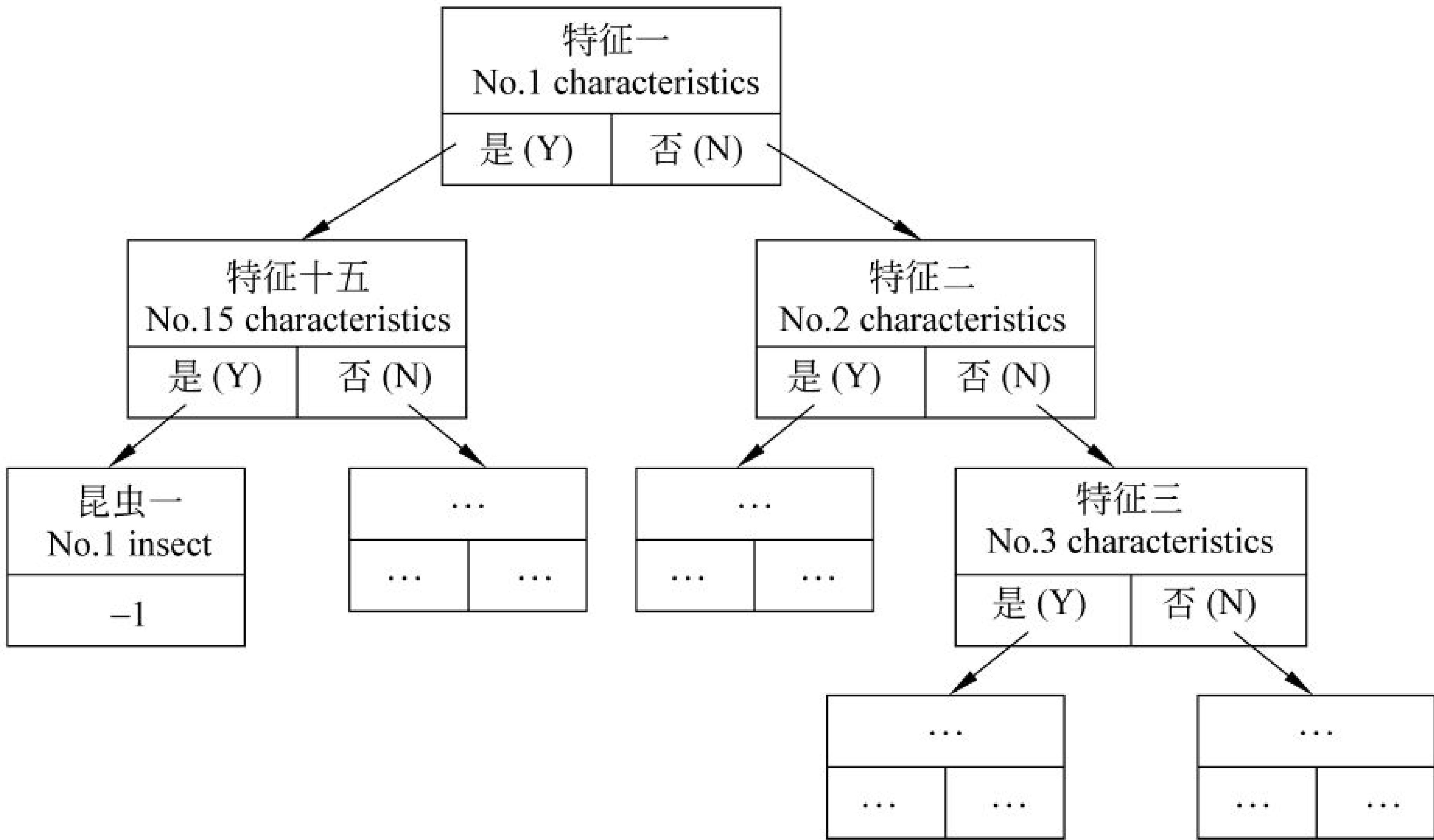


图 23-1　基于二叉树结构的农业昆虫诊断知识表示

（3）农业病虫害诊断问题求解策略与推理机制的研究。在充分考虑鱼病诊断的随机性、模糊性和不完备性基础上，结合覆盖集理论的概率模型和症状提取的模糊度，建立了基于模糊数学和覆盖集理论的诊断模型，并将该诊断模型成功用于“症状-疾病”诊断实例中。根据“疾

病-病因”诊断求解的特点以及病因事件和疾病事件之间的逻辑关系，利用覆盖集理论的节约原则构建了病因诊断的指标，将“疾病-病因”诊断问题转化为求解 0-1 整数规划的问题，并采用禁忌搜索(TS 搜索)方法来求解这一问题。通过大量鱼病诊断实例的计算结果表明，TS 搜索方法在降低误诊率和减少响应时间方面均有良好的性能。

(4) 基于案例的农业病虫害诊断问题求解策略与推理研究(见图 23-2)。基于案例的农业病虫害诊断问题求解策略与推理机制研究可以解决传统的基于规则的农业病虫害诊断专家系统存在的知识获取瓶颈和对诊断思维模拟不足的问题。基于案例的鱼病诊断推理在案例知识分析与表示的基础上，将鱼病案例转化为二值逻辑表示，进而探讨了两种不同的问题求解策略。一种是基于相似性度量模型的检索策略，另一种是基于粗糙集理论及其分析方法的检索策略。前者在改进相似性度量模型的基础上实现了一种动态的变权机制；后者通过相对约简建立多级索引来完成案例的检索。两种检索策略从不同角度处理了鱼病诊断问题中存在的不确定性。

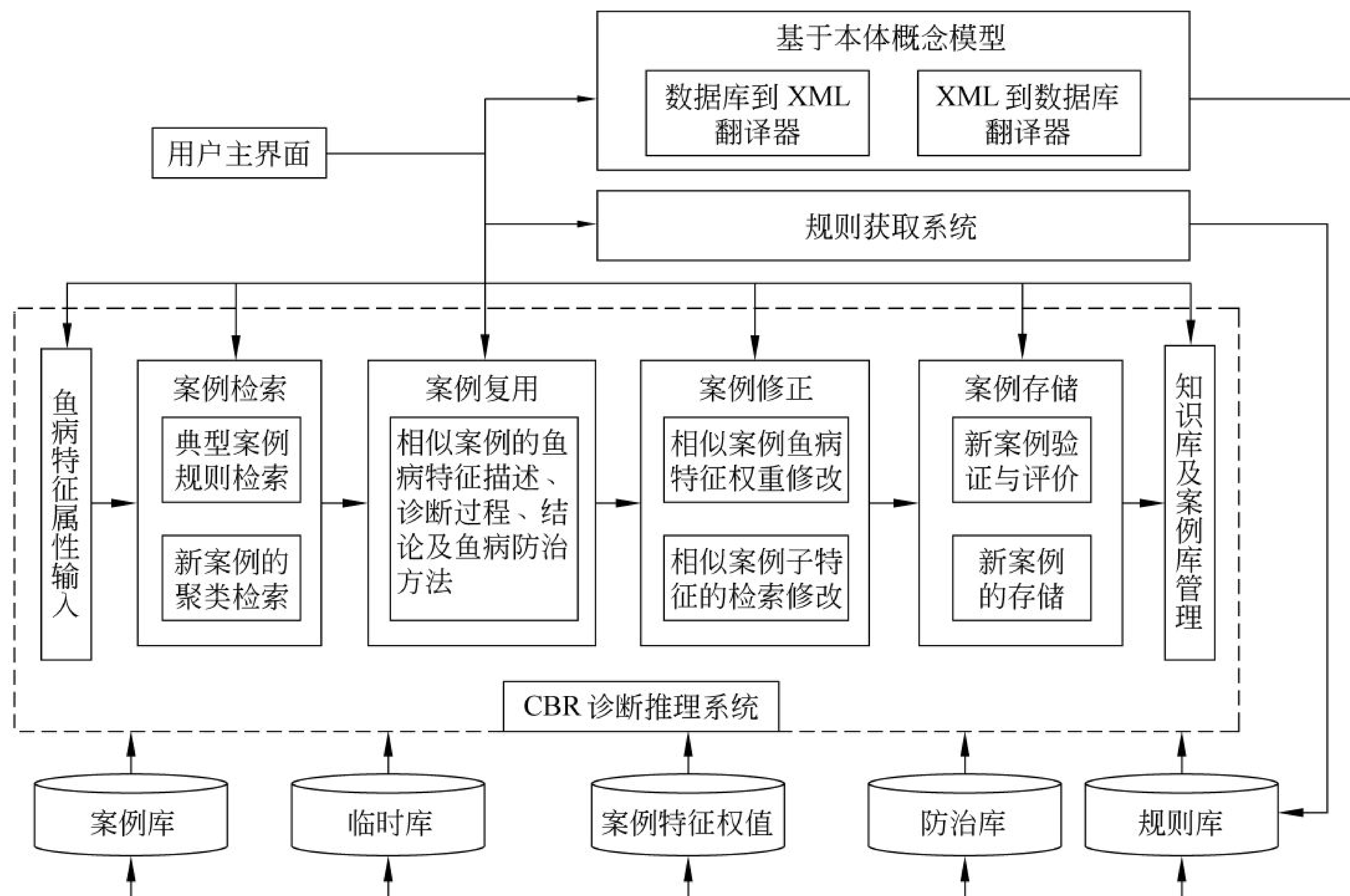


图 23-2　基于案例的推理流程

(5) 农业病虫害诊断知识获取研究。提出了基于遗传算法(GA)的知识获取模型，将鱼病诊断的知识获取问题转化为组合优化问题。并根据 AQ 算法的基本思想与学习策略，建立了一个基于整数规划的概念获取模型。实验结果表明该模型成功用于鱼病诊断知识的获取，能够有效解决系统知识的“边界效应”造成推理结果的不确定性和不可靠性等问题。

(6) 基于 B/S 模型农业病虫害诊断平台开发与技术集成。遵照 Rational 统一软件过程，采用支持面向对象的 UML 统一建模语言，研究了基于 B/S 模式农业病虫害诊断软件体系结构模型、框架模型、动态模型、过程模型、功能模型，实现了系统设计的平台无关性；采用 MVC 开发模式，在 Windows 系统平台上，采用 JSP、ASP、PHP 技术实现用户接口层，使用 Java 技

术实现业务逻辑和控制层，采用 SQL Server 2000/MYSQL 作为后台数据存储，分别建立了基于案例的、基于规则的、基于神经网络的构件，构建了农业病虫害诊断平台。实践表明，该开发方案较好地解决了构件的适应性，开发环境的无关性以及遗留工程、商业构件与新开发构件的集成问题。

(7) 基于 B/S 模型农业病虫害诊断平台应用与集成。结合示范点(北京、山东、福建、重庆、新疆)不同农作物、水产品、蔬菜、果树、家畜等病虫害的特点，对平台进行不同程度的优化与二次开发，研制了水产(鱼病、特种鱼、甲鱼)、棉花、家畜(猪、牛、羊)、蔬菜、果树等疾病诊断专家系统，并完成系统测试与初步的推广示范工作。

23.2.3　农业病虫害远程会诊平台研究

该子课题完成了农业病虫害远程会诊平台，该平台由病虫害自动诊断子系统、病虫害专家同步会诊子系统、病虫害专家异步会诊子系统构成。

(1) 系统整体设计及基本协议。系统采用 C/S 模式设计，由远程会诊服务器端(MCU)、客户端、管理员端组成。系统运行时由客户端登录服务器端建立连接，通过 MCU 进行图像与语音的分配和切换，由管理员端进行协作流程控制，使用业界普遍接受的 T.120 和 H.323 协议。T.120 用于数据协作；H.323 用于音频和视频，并支持这些协议的不同版本。

(2) 系统关键技术。采用 DirectX 技术。DirectX 技术既能保持设备无关性(开发时不必考虑支持哪种设备，只要 Windows 支持这种设备即可)，又能直接访问硬件设备(只需要通过应用程序编程接口(API)控制硬件设备，而无需了解硬件的具体细节)。

(3) 系统主要功能。

- 支持多种视频分辨率，从而满足从桌面级别到会议室应用的不同视觉效果要求。用户可根据网络带宽条件和实际会议需求，自行设定会议的视频分辨率和最大终端应用带宽，实现从桌面级别到会议室级别的不同会议需求。
- 支持多路视频：可以支持同时显示多路视频图像。在本地用户电脑负载允许的情况下，可以支持无限多路视频。
- 系统具有强大的数据协作功能使交流更加全面充分。所提供的数据协作功能包括：电子白板、文档共享、文件共享、协同浏览、桌面共享、应用共享等，并提供会诊过程中的文字交流和文件实时传输。参加病虫害会诊的人员可通过以上各种交互方式进行病虫害诊断与防治技术的支持与交流。
- 异步会诊系统：根据“病症-病名-病因-治疗”的逻辑关系，为一个格式化的 E-mail 系统开发了一个用户上传病例信息、专家根据上传病例信息进行诊断再将诊断结论和防治方法发送给用户的系统。

23.2.4　农业病虫害远程呼叫与咨询系统研究

构建了基于呼叫中心的农业病虫害远程咨询子系统，实现 Web、手机、邮件、传真等多种信息获取手段的融合，实现了用户信息库、各种数据的统计分析算法、自动呼叫分配算法以及自动语音应答，使得用户在各种情况下都能访问系统，以获取相应信息。在此基础上，为了进一步指导呼叫中心的发展，研制基于数据仓库的数据分析决策系统，实现了用户分类及信息需求预测等功能，为最终实现信息服务“个性化、针对性”提供决策支持。

(1) 呼叫咨询统一接入系统。本系统利用现代 CTI 技术，实现了对电话、传真、短信、Web、E-mail 和语音留言等多种媒体的统一受理，客户可以通过任何一种便捷的接入方式，比如一个电话号码或一个网址，就可以享受“一站式”服务，以自动服务或人工服务方式全天候地获得所需的农业信息，这种方式极大地拓展了农业信息的传播时空距离。本系统以其便捷的接入方式、人性化的服务和高效的客户信息管理模式，为解决农业信息化的“最后一公里”和“信息孤岛”问题提供了技术支持，大大缓解了农业信息服务机构在客户服务方面存在的与客户交互性差的矛盾，为农业生产和经营提供了信息支撑。

- 自动服务：通过自动应答系统(IVR)，客户可以通过电话按键选择的方式来获得录制好的语音农业信息，在自动服务无法满足的时候，可以转接人工服务或者语音留言。
- 人工服务：人工坐席可以接受客户的电话、传真、短信、E-mail 和 Web 方式的服务请求，坐席员可以通过语音、短信、传真、E-mail、文本交谈和同步浏览的方式满足客户的呼叫请求，从而更加人性化地回答用户提出的问题，并且坐席员可以通过系统挂接的农业信息数据库查询用户所需的信息。

(2) 呼叫中心数据仓库构建。针对呼叫咨询系统存在的“数据丰富，信息匮乏”问题，提出了利用数据仓库的方法来解决农业呼叫中心的数据分析问题。即以数据仓库为基础，以联机分析处理和数据挖掘为手段的解决方案。研究结果表明，本研究的开展使得决策分析人员能够利用数据仓库中的数据进行更科学、更理性的决策，为改进呼叫中心的服务提供决策支持。这种方法改变了传统的数据分析手段，提高了数据的分析和应用水平。

(3) 客户信息数据挖掘模型与系统。借鉴客户关系管理(CRM)思想，采用数据挖掘的方法对呼叫中心的客户信息进行深层次分析，重点采用决策树方法进行客户细分和信息需求分类，生成决策树，提取规则集，完成知识发现过程。大量客户数据的实证分析表明，数据挖掘方法可以发现隐含的客户信息需求，这为以后有针对性地提供个性化的农业信息服务提供了决策支持。

23.2.5 移动式农业病虫害诊断决策系统研究

设计了基于移动 Web 服务和智能手持设备的移动式农业病虫害诊断决策系统平台；针对鱼病与猪病专家系统实现了其知识库、数据库和推理机的改造和优化，成功地移植到 PDA 手机中，设计并实现了基于嵌入 Java 技术的移动式鱼病与猪病诊断专家系统，已完成实际环境测试。系统可以广泛应用于摩托罗拉、诺基亚、索尼爱立信等品牌的中高端手机中。

(1) 系统整体结构设计。该系统可以提供 B/S、C/S、本地(本机嵌入式系统)三种访问移动专家系统的模式，其体系结构如图 23-3 所示。移动式专家系统由服务器端子系统、客户端子系统和网络通信服务系统组成，可以满足用户在有稳定的无线网络连接环境和无稳定的无线网络连接环境情况下现场诊断的需求。

(2) 服务器端子系统。服务器端子系统由知识获取系统、服务器端专家系统和移动 Web 服务器组成。

- 知识获取系统。将神经网络、粗糙集、蚁群算法等大规模智能计算技术用于知识获取，在算法实现上采用效率较高的 C/C++ 语言，鱼病案例数据和相关辅助与结果数据以 SQL Server 数据库存储。



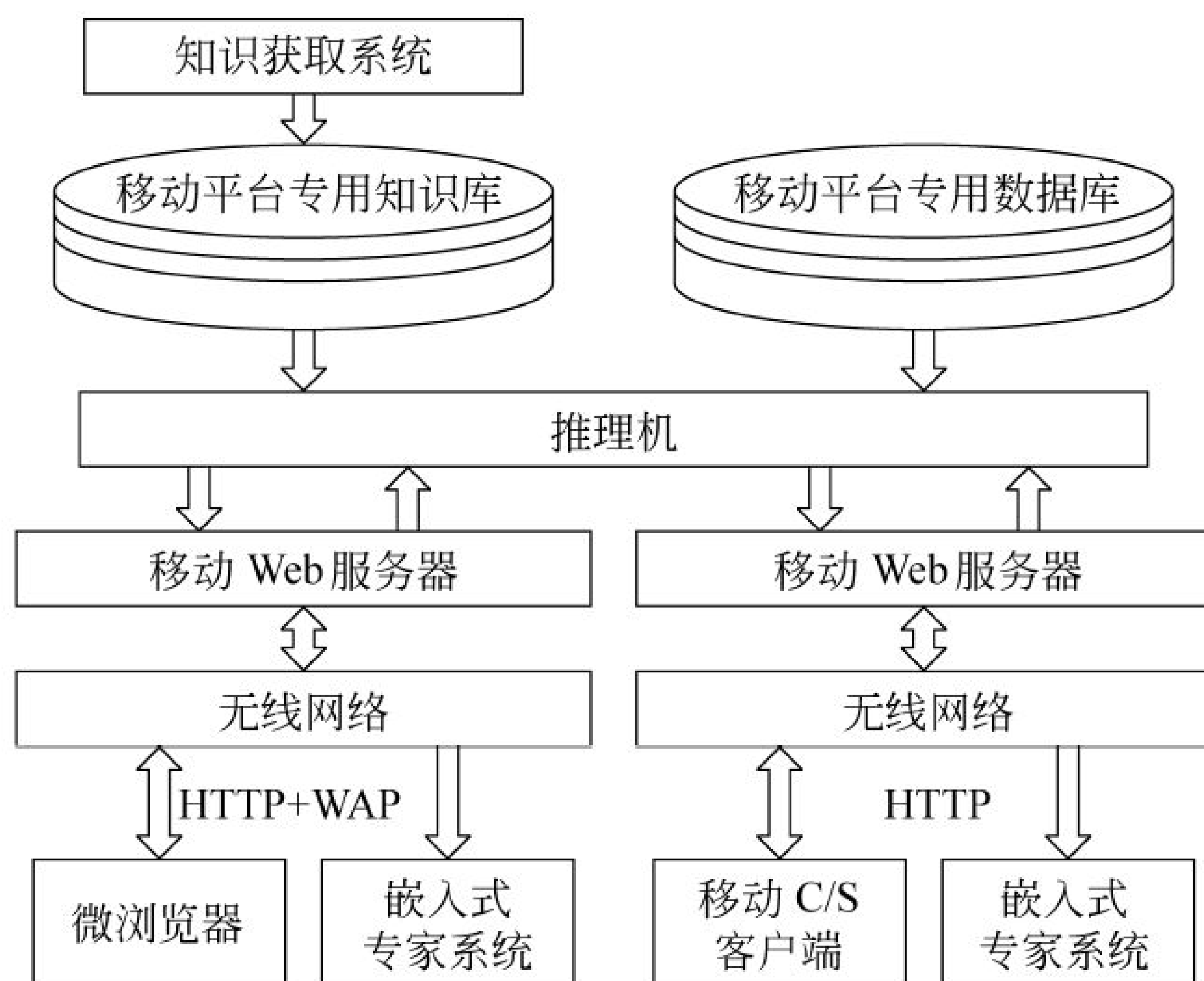


图 23-3　系统体系结构

- 服务器端专家系统。服务器端专家系统由移动专家系统专用知识库、数据库和推理机组成。移动平台专用知识库中的数据由知识获取系统提供和更新，移动平台专用数据库存储着移动专家系统的公用数据（如事实数据、资料数据、中间结果数据等），存储在 SQL Server 数据库服务器中。其推理机用 Java 中间件实现。
- 移动 Web 服务器。移动 Web 服务器提供用户访问界面，以 Apache＋Tomcat 作为 Web 服务器，以 SQL Server 作为后端数据库服务器，用户界面采用 JSP 和 Servlet 技术实现。

（3）客户端子系统。客户端子系统包括三种类型的客户端，即嵌入式鱼病诊断专家系统、微浏览器型客户端和移动 C/S 型客户端（如图 23-3 所示）。

- 嵌入式鱼病诊断专家系统。该嵌入式移动专家系统是一个完整的本机专家系统，通过在移动计算设备（如智能手机、PDA 等）中嵌入知识库、数据库和推理机等实现，可以通过无线网络下载、更新。支持随时随地的现场诊断，使用中不受网络环境的影响。该系统采用 J2ME 技术实现，具有良好的平台适应性和移植性。其知识库、推理机通过服务器端子系统的知识获取与推理机制产生，因此具有与服务器端相当的专家推理能力。系统数据存储采用 J2ME 特有的“记录管理系统（RMS）”数据格式和文本数据库格式混合存储方式，既保证了存储效率又满足数据更新的需要。系统通过移动设备的 Java 应用管理器（JAM）可以通过无线网络下载和更新，保持与服务器端系统版本的一致，从而不断提高系统性能。
- 移动 C/S 型客户端。移动 C/S 型客户端应用模式是针对具有无线网络支持但不支持微浏览器的手持移动设备专门设计的。它实际上包括两个主要的组成部分：移动客户端软件和服务器端专家系统服务。移动客户端软件，提供用户访问后端服务器专家系统的功能并提供使用界面，用 J2ME 的 MIDlet 小程序实现，只要支持 J2ME（KJava）的手持设备都可以使用；服务器端专家系统服务采用 JSP 和 Servlet 技术实现，它们接收客户端发来的服务请求，收集专家系统推理诊断所需的由用户提供的证据信息，调用专家系统的数据库、知识库和推理机进行推理，并负责向客户端返回推理所得的诊疗结论和解释信息。客户端 MIDlet 程序通过 HTTP 协议访问服务器端 JSP

和 Servlet 程序提供的相关服务，并负责与其交互。

- 微浏览器型客户端。微浏览器型客户端应用模式是针对具有无线网络支持并带有微浏览器的手持移动设备设计的。它包括两部分：移动客户端微浏览器软件和服务器端专家系统服务。移动客户端微浏览器软件由移动设备本身提供，是设备自带的访问无线因特网的软件工具，其功能和使用方式与普通 Internet 浏览器相近，只是受屏幕尺寸限制，界面功能不如普通 Internet 浏览器，只要设备支持无线网络和 WAP 协议就可以访问无线因特网服务。其服务器端专家系统服务功能与移动 C/S 型客户端应用模式的服务器功能类似，只是使用的协议和服务模式不同而已，即这里使用的是 HTTP 和 WAP 协议，以无线 WWW 服务为主。微浏览器模式的服务器端实际上是一个 WAP 网站系统，用 JSP、Servlet 和 WML 技术实现，JSP 和 Servlet 服务程序不但负责专家系统的调用，还要为客户端提供界面服务，即以 WML 文件格式将有关信息返回给客户端微浏览器。

23.2.6 基于 GIS 的农业病虫害广域分析

以山东省烟台市为主要示范区，设计开发了“基于 ArcObject 病虫害地理信息系统”后台管理软件、“基于 MapObjects 的果树病虫害地理信息系统”单用户管理软件和“基于 WebGIS 果树病虫害信息发布系统”前台网络管理软件。该平台根据烟台地区的具体情况和果农的要求，采用目前先进的 GIS 技术，为用户提供病虫害信息查询和图形操作功能，实现了对作物种植区域的空间信息、属性信息和多媒体信息的综合管理，直观地表现了病虫害的分布特点、发病特征、危害程度及其未来数天或数月内的动态变化趋势。系统通过网络共享可以为更多的各级管理、规划、决策人员提供直观综合的决策信息。

(1) 系统总体设计方案。根据不同的用户需求和各级管理职能的需要，针对果树病虫害的特点并借鉴其他 GIS 系统的经验，将系统用户分为管理员和普通用户两个层次。子系统总体设计为三大部分。

- 服务器端采用 ArcSDE 软件，在关系型数据库 ORACLE 的基础上实现海量空间数据的数据库高效管理。基于 ArcObjects 开发的系统只允许管理员进行数据预处理操作，管理员更多的是扮演数据组织、管理、分析的角色，在系统中起着最重要且不可缺少的作用。普通客户端分为两部分，一是基于 MO 的应用程序，可以实现统计、查询、诊断和预测等功能，为市县级（地区）相关部门的辅助决策提供帮助。二是采用 ArcIMS 软件，可以实现地理信息的网上发布，可使用户通过 IE 浏览器就可以看到自己想要的地理信息。系统配置方案如图 23-4 所示。
- 系统软硬件配置：操作系统在服务器端选择 Windows2000 Server 或 Windows NT，客户机端选择 Windows9x/Windows2000 Professional/Windows XP 等；数据库管理系统选定 Oracle9i；开发工具包括 ArcGIS 桌面软件、MapObjects、ArcObjects、ArcIMS、ArcSDE 和 MS Visual Basic6.0 等；硬件设备包括服务器、网络交换机、客户机和输出设备(彩色绘图仪、激光打印机)等。

(2) 基于 ArcObjects 的烟台苹果病虫害地理信息系统。本系统属于网络地理信息系统的后台管理系统，负责系统的数据管理、空间分析、基于神经网络的预测预报、病虫害诊断等，并将分析处理的结果通过 SDE 存储到 Oracle 数据库中，供网络用户端查询使用。本系统采用 Vb6.0＋ArcSDE＋ArcObjects＋Oracle9i 的开发方式，系统开发平台 Win2000Server。



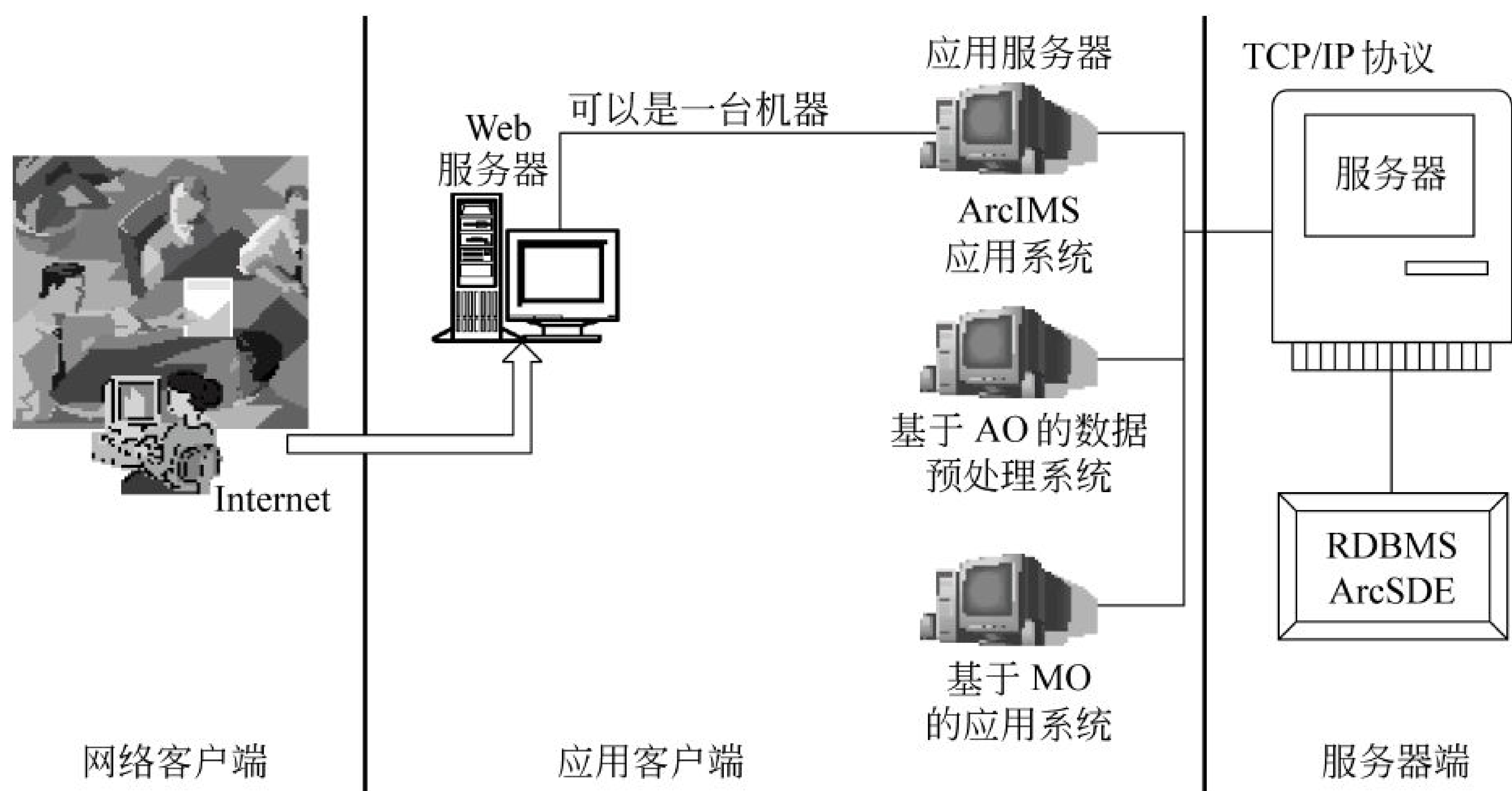


图 23-4 系统配置方案图

Vb6.0 开发系统具有简单、快速的优点，而且使用它可以方便地调用 ArcObjects 二次开发组件。ArcSDE 是空间数据引擎，通过它可以方便地将空间数据导入到 oracle9i 数据库中，ArcObject 是 ESRI 公司提供的组件库，也是 ESRI 公司的核心产品，功能非常强大，系统所需要的各种 GIS 操作主要由它来完成。Oracle9i 是企业级数据库，存储数据量大、吞吐量大、数据安全、满足病虫害数据存储量大、数据的频繁访问以及数据扩充的需要。

(3) 基于 MapObjects 的果树病虫害地理信息系统。本系统是客户端应用程序，利用面向对象的知识表示方法、Visual Basic 语言编程、人工智能技术、决策支持系统理论，实现了知识库、模型库、推理机与地理信息库的有机动态结合，动态显示病虫害的灾情、险情地理分布图，直观性强，并且利于掌握病虫害的全局地域分布规律，进行有效防治，对作物病虫害的发生趋势做出准确预报，为农业生产者提供科学的防治决策。主要功能包括数据库管理、历史信息查询、果树病虫害信息查询与诊断、数据分析处理、基于神经网络的预测预报、信息输出等。

(4) 基于 WebGIS 的果树病虫害信息发布系统。本系统是客户端应用程序，采用用户登录权限管理，将用户分为普通用户和管理员用户，普通用户享有除数据维护外系统提供的所有功能，可以进行病虫害信息的浏览、空间查询、统计分析以及诊断防治、蔓延预测等，而管理员用户除了享有普通用户的权限外，还具有使用系统的总体数据维护、数据录入和更新等功能的权限。

(5) 病虫害预测预报的研究。以山楂叶螨为研究对象，利用 Matlab 提供的神经网络工具箱中的功能在后台完成创建结构为 4-15-5 的 BP 网络，采用 L-M 优化算法(trainlm)训练网络。BP 神经网络与缓冲区分析相结合实现预测结果的显示。根据某地区的平均气温、最低气温、日照时数和降雨量四个参数，BP 网络的预测结果分不同颜色表示灾害的严重程度，同时依据风向和风力确定缓冲的位移量和缓冲距离。

23.2.7 虫害远程自动识别与诊断平台

虫害远程自动识别与诊断平台研究了病虫害图像获取软硬件一体化平台，定义了用于虫害自动识别的特征参数，改进了虫害自动识别的分类方法，研究了客户端与服务器端进行远程识别的交互方法，完成了基于计算机视觉的虫害远程自动识别系统的设计和开发工作。详述如下。

(1) 病虫害图像获取软硬件一体化平台。虫害的图像数据是虫害自动识别的主要依据。图像获取的非标准化直接关系到自动识别的准确率与可行性。因此设计统一的图像获取平台才能保持获取图像数据的一致性,才能使自动识别算法具有稳定性和实际意义。基于此,研制了一套包括数字摄像机、工业镜头、光源、图像支架、计算机为一体的图像获取平台专门用于昆虫图像的获取。与虫害自动鉴定软件的客户端配合使用,用于在计算机上实时从图像平台上的摄像机获取昆虫图像并用于自动鉴定。

(2) 用于虫害自动识别的特征参数。对于昆虫自动识别来说这是一个具有挑战性的课题,除了我们项目组外很少有科学家进行这方面的研究,因此少有相关参考文献等技术资料。为了稳定地实现基于图像的昆虫自动识别,我们搜集了许多虫体个体的实际情况并结合计算机视觉理论提出了一套用于昆虫自动识别的新的特征系统:包括基于数学形态学的特征、基于颜色的特征和基于纹理的特征相结合的综合识别特征。

(3) 虫害自动识别的分类方法。由于有新的虫害特征加入,我们发展了虫害自动识别的分类方法,将以前的二叉分类识别方法上升为基于神经网络的模式识别方法。这在昆虫自动识别领域是一创新之举。

(4) 客户端与服务器端进行远程识别的交互方法。为了实现真正的远程自动识别,我们将识别系统从单机运行环境提升至基于 TCP/IP 协议的网络系统。网络系统真正实现了客户端与服务器端的分离,服务器端对来自客户端的自动识别请求进行处理,检索特征数据库,进行分类识别,并将最终的识别结果返回给客户端。

(5) 基于计算机视觉的虫害远程自动识别系统。基于上述研究,采用 Visual C++ 编程语言设计了一个初步的虫害远程自动识别系统。该系统的客户端可以实时地从图像获取平台上获取昆虫图像,并进行特征提取。客户端将提取的特征通过 TCP/IP 协议传给连接在网络上的远端服务器,远端服务器基于特征数据库进行自动识别,并将识别结果通过 TCP/IP 协议返回给客户端。实践结果表明系统达到了较好的识别效果,这在昆虫远程自动识别上是一大进步。

23.3 项目采取的技术路线和关键技术

23.3.1 研究方法与技术路线

本项目是一个复杂的系统工程,涉及植物保护、动物疾病学、昆虫学、农艺学、植物学、数学建模方法、优化理论、计算机软件设计、计算机通信、人工智能、数据挖掘等多学科多领域的知识内容。

研究的总体目标是提高农业病虫害防治能力,实现农业可持续发展;集成知识工程、图像识别、GIS、网络多媒体等新技术,研制开发网络化农业远程诊断技术平台与“以点带面”、“点面结合”的系统集成与体系;为国家在大尺度空间内的病虫害防治系统奠定基础。根据这个研究目标,相应的系统技术路线如图 23-5 所示。

关于系统技术路线的详述如下:

第一步,利用环境监测、GIS/GPS 技术对影响农业病虫害诊断的关键因子进行监测与研究,建立相应的各种模型,进而形成包含 5 个相辅相成、相互依赖、相互补充的子系统所构成的农业病虫害诊断数字化系统。该系统能够“从点到面,从微观到宏观”辅助生产管理部门、动物



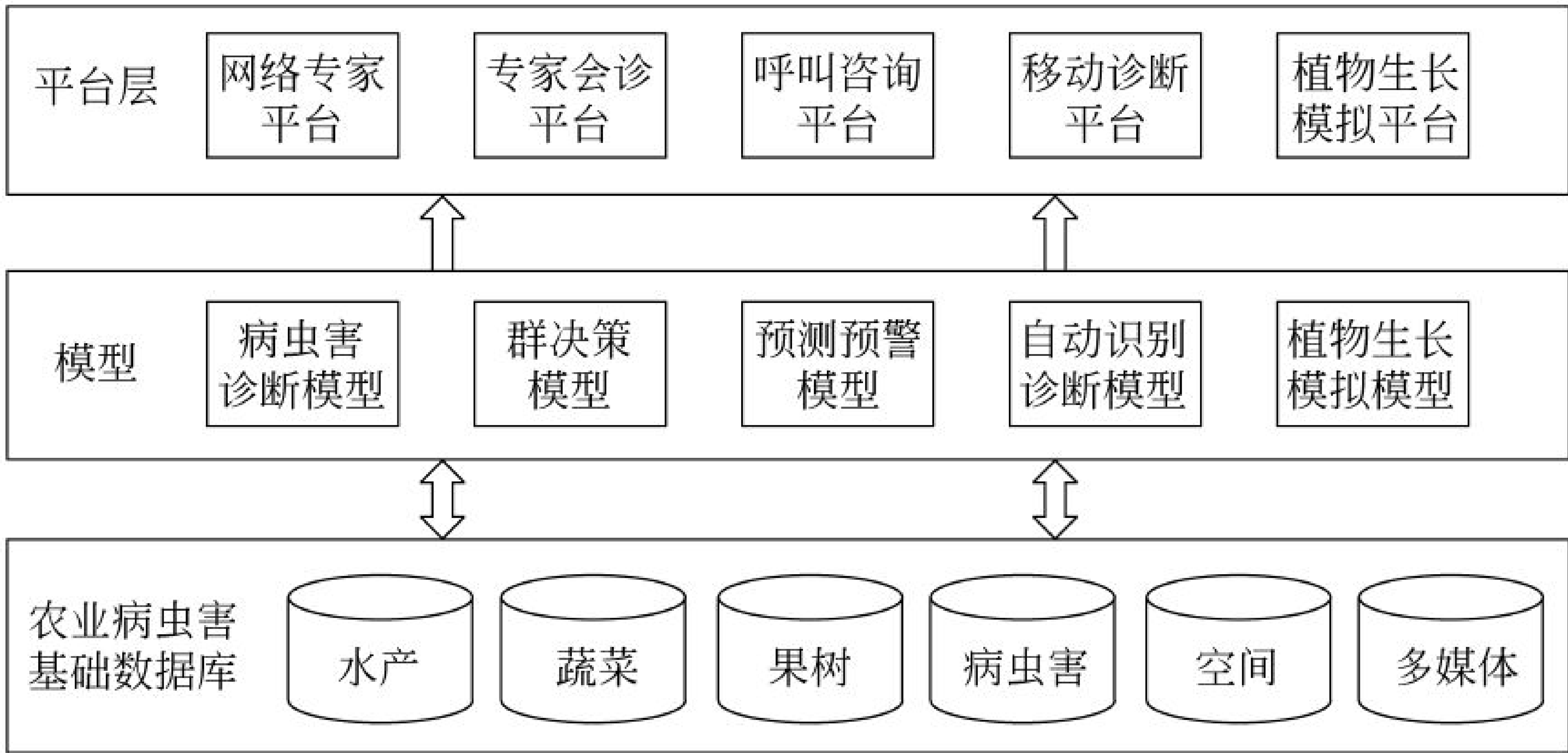


图 23-5　系统技术路线图

医院等进行农业防疫和病虫害诊断。

第二步，在全面细致分析的基础上，对课题的各个方面进行科学规划与组织，并行交叉地开展工作，在对用户需求进行分析的基础上，通过网上咨询和农业专家访谈，系统分析和收集整理相关知识，建立农业病虫害问题的表示模型和求解模型，建立基于 B/S 模型的专家系统。

第三步，运用异地信息交互传输技术和计算机协同工作的思想，利用网关服务器、ILS 等设备和软件将远程语音、视频、数据数字化，采用 NmSDK、IE SDK 等网络及多媒体软件技术，设计农业专家和用户能够进行语言、图像、文字、文件实时交互的“面对面”专家会诊系统。

第四步，研究将集成组件技术、远程诊断与呼叫技术用于鱼、棉花与蔬菜、果树病虫害的远程诊疗系统，重点解决专家系统模式、坐席工作人员模式和坐席专家模式的农业病虫害预报、诊断和防治。

第五步，根据农业病虫害的特征和规律，结合空间信息、地理特征、气象特征建立基于 GIS 技术的数字化农业病虫害诊断系统。

第六步，采用基于 PDA 的移动式农业病虫害诊断系统。

23.3.2　关键技术研究和突破

1. 农业病虫害领域知识表示模型

- 运用领域工程思想，构建了农业病虫害诊断领域知识表示模型与形式化体系。该模型的建立有效提升了农业病虫害诊断平台的可重用性。
- 结合农业病虫害（以鱼病专家为例）的诊断思维模式，综合鱼病诊断专家系统各案例元素间的关系，给出了基于案例推理的鱼病诊断概念模型，特别是针对鱼病案例属性定性、离散化的特点，将鱼病案例转化为二值逻辑表示，建立了基于相似性度量模型和基于粗分析方法的两种案例检索策略。前者在改进相似性度量模型的基础上实现了一种动态的变权机制；后者通过相对约简建立多级索引来完成案例的检索。两种检索策略从不同角度处理了鱼病诊断问题中存在的不确定性。
- 提出了基于遗传算法（GA）的知识获取模型，首次将其应用在鱼病诊断推理系统中，用于获取鱼病诊断规则。实验结果表明该方法能够有效解决系统知识的“边界效应”造成推理结果的不确定性和不可靠性等问题。

- 提出处理多种不确定性诊断知识共存的方法。针对专家经验知识具有随机性、模糊性、未确知性等特点，在节约覆盖集理论及其概率模型的基础上，结合模糊数学构造了合理的处理这些不确定性知识表示和推理的模型，并采用相应的序贯算法进行求解。

2. 农业病虫害诊断与推理算法

- 发展了原 Bugvisux 系统中虫害自动识别的分类方法，将以前的二叉分类识别方法上升为基于神经网络的模式识别方法。
- 针对案例特征权值向量难以确定的问题，提出将遗传算法(GA)应用于特征权值优化来提高检索质量；针对现有的 RBR 推理和 CBR 推理的优缺点，提出了将 RBR 与 CBR 推理相结合的层次聚类方法来缩小鱼病诊断案例的搜索空间，提高系统的检索效率。
- 提出了"症状-疾病-病因"的双层因果诊断模型及将病因诊断问题转化为优化问题求解的新思路。在分析鱼病发生和传播机理的基础上，提出了双层因果诊断模型，并对鱼病诊断问题进行了形式化描述，为鱼病诊断系统的实用性、针对性提供了切实的理论保障。利用覆盖集理论构建了病因诊断的指标，开创性地提出一种将病因诊断问题描述为优化问题，采用现代组合优化方法求解诊断问题的新思路，并采用大量鱼病诊断实例证明该算法的有效性。研究将有助于提高鱼病诊断的准确率和速度，是专家系统借助于优化算法的一次尝试。

3. 数字植物生成模型

- 提出了 GreenLab(青园)模型植物建模方法，这是一个离散的、基于植物学原理的，包括植物形态结构变化以及生态生理过程的动态随机模型系统。相关建模方法给出了新的植物生长建模理论框架。
- 首次提出并发展了双尺度自动机模型。该模型以自动机与随机过程为基本理论方法。同时，根据植物特殊的树状结构提出了特有的模型方法。在对植物拓扑结构的分析上，通过对芽的行为(活动、休眠、死亡、分支)的数学模拟，建立了以结构单元(growth unit)对应于宏状态，叶元(metamer)对应于微状态的形态发生模型。这两种形式的状态构造了双尺度自动机模型。双尺度自动机的特有结构是"基于分层上的串行结构"，并且以状态转移图的形式表述结构的随机变化规律。
- 首次提出了子结构方法。在"子结构"研究中，我们不仅推导了确定性条件下植物器官计算的显示表达结果，同时，还推导了"随机子结构"计算的显示表达模型，由此建立了植物器官发生数的均值与方差的计算方法，并给出了稳定环境中植物生物量生产的统计特性的近似计算方法。特别是"随机子结构"计算模型，为通过分析解方式探讨植物生长规律提供了可能，而传统方法只能是通过大量仿真结果的数据统计得到相关分析的观察。模拟比较结果完全证实了新方法的优越性。此外，"子结构"也为在植物生长可视化中合理、有效地利用"层次细节"技术提供了前提。
- 首次将模型估计引入这一问题的研究，突破了传统方法置该研究停留于统计数据的分析。通过对实验数据的分析和模型参数的估计处理，我们成功地重建了上述植物生长与结构的动态发展过程。这一研究的重要意义在于为模型应用于产量预测打下了必



要基础。同时，它也使考察模型参数变化与作物产量的理论或数值分析成为可能。

4. 平台开发与集成技术

- 研制了一套包括数字摄像机、工业镜头、光源、图像支架、计算机为一体的图像获取平台，专门用于昆虫图像的获取。并研制了一套从上述设备实时获取图像的软件装置，用于在计算机上实时地从图像平台上的摄像机获取昆虫图像，以作自动鉴定之用。
- 以多点控制单元（MCU）为中心，采用包交换方式、以 H.323 协议为标准的音视频通信协议及以 T.120 协议为标准的数据通信协议实现了远程交互会诊平台。
- 采用本体论的基本理论和方法，提出了基于领域本体的农业病虫害诊断系统集成的概念体系，针对翻译器的构造，以 XML 与数据库之间的转换为例给出了具体算法，有效地解决了不同平台与系统集成时的知识共享及二义性问题。

5. 农业病虫害广域分析

- 提出 BP 神经网络与地理信息系统相结合实现农业病虫害预测预报。运用 BP 网络模型对农作物病虫害的发生量进行预测预报，与现有的判别方法相比，我们的方法具有通用性、客观性和科学性的特点；GIS 技术使得病虫害的灾害程度在电子地图中形象直观地显示出来。此研究为建立预测模型库奠定了基础。
- 将地理空间数据库与农业病虫害的诊断知识库、标准图像库和诊断模型、农业害虫的数字化特征库、病虫害知识库等有效动态集成，通过 WEBGIS 快速查询、匹配准确。

23.4　项目推广和示范的社会效益

23.4.1　项目的示范基地建设与企业合作

课题组充分感受到：作为科研单位，研究的内容既要考虑技术的先进性，又要兼顾基层单位的实际生产需求，走“产、学、研”相结合的道路。应该尽快把先进实用的科研成果转化为生产力，转化为经济效益、生态效益和社会效益，以降低劳动成本，增产增效，提高农民收入，改善农民生活质量，促进农村产业结构调整，提升农村经济的竞争力，使农村经济逐步迈入可持续发展的新轨道。为了提高项目成果研究的科学性与技术性，以及使项目研究成果及时得到应用与示范，课题组的工作主要从以下两方面展开。

1. 基地建设

根据项目合作伙伴的变更以及进展的实际情况，选择了北京、天津、山东、福建、重庆与新疆 6 省市，根据省市各区、县、乡镇农业病虫害的实际发生状况以及现有网络技术，经课题组充分讨论研究，将技术研究与推广培训相结合；建立由地方政府领导及地方部门领导组成的行政领导体系，有关专家及地方科技人员组成的技术领导体系，县、乡、村、户四级推广体系，分别负责项目组织、技术培训、改进和推广实施；从中心示范区、带动面和辐射推广面不同层次，分层实施该项目，使平台应用不断推广深化，并取得显著经济效益，提高了农民收入，促进了当地农业结构的调整。

为充分发挥信息技术在农业科技推广中的作用，在网络基础薄弱区域，与相关基础单位合

作，课题组购置了服务器及相应网络设备，在示范点技术推广站建立了农业病虫害信息中心和分中心，部分当地政府也积极配置相应经费，为示范户购买了计算机和调制解调器等网络设备，建立了县、乡、户三级网络体系，并由专人负责系统的运行维护及资料收集、整理和更新。课题组成员还对计算机使用人员进行了专门培训。这些基础工作不仅在项目中起到数据提供者的作用，还作为中心示范区为其他示范区提供了范式。以下是部分具有特色的关键推广与示范区：

- 与山东烟台市农业信息中心协作，制定烟台市果树病虫害诊断、预测预报等试验示范实施方案。在本课题的研究过程中，作为开放实验室，与烟台市农业信息中心签订了合作合同，旨在烟台市开展果树病虫害广域分析试验与示范，提高烟台市果树病虫害管理水平。示范区按照市、县、乡管理单元，以烟台市农业信息中心作为系统管理中心，在烟台市选择两个县，在每个县选择两个乡。烟台市提供了烟台市果树种植分布图；烟台市行政区划地图；烟台市土地利用图（最小区域为乡镇，示范点可设置在村）；烟台市果树种植及为病虫害管理等发明的特殊数据与特殊管理知识；烟台市历史气象数据和实时气象数据。这些内容为本系统的开发和试运行奠定了基础。
- 与山东东营市科技局合作，制定了东营市淡水养殖水产品疾病诊断等试验示范实施方案，以提高东营市淡水养殖水产品病虫害管理水平。示范区按照市、县、乡管理单元，以东营市垦利县农业信息中心作为系统管理中心，选择 14 个乡。东营市垦利县海洋水产局提供养殖水产品病虫害信息以及相关养殖环境等具体资料与信息，还有为病虫害管理等发明的特殊数据与特殊管理知识。这些内容为本系统的开发和试运行奠定了基础。
- 与天津农学院、天津水产网、天津水产养殖场等合作，针对天津水产的实际情况，重点研究了鱼病病因，特别是从水质角度出发研究了水体环境对鱼病产生的影响，形成了不同于北京市鱼病诊断专家系统的版本；并在相关单位的协助和当地政府的支持下，根据天津市各区、县、乡镇水产养殖的基础、条件和水平，中心示范区选择在养殖规模大，区域相对集中，基础和条件相对较好的地方，课题组在每个核心示范区选择了3～5 家示范户，在项目实施初期免费为其安装系统，并在技术指导方面进行了重点扶持。在示范户的带动下，技术得到了大规模普及。
- 在福建省农业科学研究院植保所建立了包括农业病虫害基础信息咨询服务系统平台、农业害虫自动识别与诊断平台、农业病虫害辅助诊断平台及农业病虫害远程会诊平台在内的农业病虫害远程诊断网，并与福建省龙海市、福建省南平建瓯市当地科技局达成应用合作协议，在当地设立病虫害远程诊断平台应用示范点，以本项目的示范网站农业病虫害远程诊断网为依托，在市农业学校建立了呼叫服务中心，为当地提供多方位的植保信息咨询服务。同时，与重庆渝北国家农业园区开展了番茄病害远程辅助诊断应用等方面的试验性合作。

2. 企业合作

课题组在平台选择、开发总体方案的制定以及软件开发过程中，与北京坤泰伟业科技有限公司、ArcGIS 中国培训中心、华康科技开发公司、北京佑格科技发展公司等建立了密切的合作关系，并进行了多次技术交流。这种跨部门、跨领域的合作有效地支撑了研究课题的顺利进行。以下是部分具有特色的关键合作企业：

- 北京蒲公英信息技术有限公司。这是一个专门从事网络地理信息系统的公司，具有丰富的开发经验和开发实力。该公司为本课题开发过程中遇到的各种技术问题提供解决方案，辅助完成了整个系统的解决方案和总体设计，保证了整个项目的进度安排和质量。
- 北京坤泰伟业科技有限公司和北京佑格科技发展公司。这是两家专门从事农业专家系统与决策支持系统开发的公司，具有雄厚的开发实力、丰富的推广营销经验，同时还拥有大量相关客户与农业信息资料。他们为本课题开发过程中遇到的各种技术问题提供解决方案，辅助完成了整个系统的解决方案和总体设计以及系统推广与示范的策划，保证了整个项目的进度安排和质量。
- 福州易达信软件科技公司。由项目参加单位福建省农科院与该公司合作进行本项中农业病虫害远程会诊平台的开发与研制，在技术上对系统中所采用的H263++视频压缩标准及传输协议进行改进，大大提高了远程会诊平台的性能。
- ArcGIS中国培训中心。主要对课题组成员进行最新的ArcGIS合作培训，让课题组成员更快、更好地掌握新版ARCGIS软件，使系统平稳升级到新版本。

23.4.2　其他经济、社会效益分析与评述

以计算机网络技术、多媒体技术和通信技术为支撑的农业病虫害远程诊断平台不仅拥有高层次、多功能的农业专家知识，并能模仿人类的推理过程，以形象、直观的方式向使用者提供解决各种养殖问题的咨询服务与决策方案，是用于解决农业病虫害防治中诸多问题的有效途径，可以普及农业病虫害科技知识，提升农业防灾与抗灾的整体水平。

(1) 有效地提高农业病虫害防治能力，为进一步提升预警能力提供了基本资料保障。技术平台的建立将农业病虫害的监测技术、诊断技术与防治技术有机地结合起来，并且利用了先进的遥感遥测系统、全球定位系统、计算机网络通信技术、人工智能技术、专家系统和决策支持系统技术、地理信息系统、多媒体技术等对重大农业生物灾害的发生监测、诊断，并据此进行防治决策，为进一步进行预警研究提供了良好的平台，以达到灾害性生物发生、危害等动态始终处于人类掌握和控制之中的最终目的。

(2) 加速农业灾害控制科技的研究与推广，解决“产学研”脱节的问题。技术平台的建立有效地遵循用户导向的指导思想，加强科研成果与推广体系的有效整合，有助于实现教学、科研和生产三结合，加强农业灾害科技推广和科学普及宣传工作，提高全民族的防灾减灾意识和水平。

(3) 加速形成技术标准，实现技术的共享，推进农业灾害控制科技革命。技术平台的建立势必要加快重大农业生物灾害监测、预警与防治技术的整合以及推广体系的网络化进程，促进农业生物灾害技术规范的形成和有效共享，并且通过“强强”联合、连锁服务和网络化运行，降低防治服务成本，增强竞争优势和生命力，进一步推进农业灾害控制科技革命。

23.5　项目实施的启示

项目实施给出了以下启示：

(1) 目标管理与过程管理相结合。根据课题的目标任务，强化目标管理，制定科学合理的考核目标，制定责任制；充分发挥各子题部门和地方合作单位的作用，加强实施过程中的规范

管理和动态检查，确保工作质量。

(2) 加强子题之间的合作、协作与交流。通过例会、月报、年中报告会等形式进一步加强子题之间的相互学习、交流和监督，保证子题按计划进行。

(3) 强化课题研究与示范区建设的有机结合，加强开发系统的实验与示范工作。技术创新和产品研发都围绕相关的示范区开展研究与示范工作，保证所开发的产品有明确的社会需求，同时也为示范区的综合示范提供技术支撑。

(4) 加强人才队伍建设。为实现培养年轻人尽快脱颖而出，课题组非常注意年轻学术骨干的培养，大胆放手让青年骨干承担重要任务，努力加强人才培养和队伍建设，逐步建立一支学科分布合理、技术领先、团结协作、有实力、能攻坚的精干科技队伍。

(5) 加强宣传普及工作。利用网络、新闻媒体等多种手段进行宣传报道，营造社会氛围，使公众认识和接受现代信息技术在农业病虫害防治中的重要意义，凝聚社会各界力量，确保现代农业技术研究可持续发展。

(6) 充分利用国际合作渠道，开展国际交流。通过与世界上在该领域活跃和领先的研究人员合作，我们不仅缩短了人员培训过程，同时使中国的学术研究很快进入国际前沿，并取得了许多实质性的学术进展。